



US 2009004766A1

(19) **United States**

(12) **Patent Application Publication**
Lofton-Day et al.

(10) **Pub. No.: US 2009/004766 A1**

(43) **Pub. Date: Feb. 19, 2009**

(54) **METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF COLON PROLIFERATIVE DISORDERS**

(75) Inventors: **Catherine Lofton-Day**, Seattle, WA (US); **Matthias Ebert**, Muenchen (DE)

Correspondence Address:
DAVIS WRIGHT TREMAINE, LLP/Seattle
1201 Third Avenue, Suite 2200
SEATTLE, WA 98101-3045 (US)

(73) Assignee: **Epigenomics AG**, Berlin (DE)

(21) Appl. No.: **10/562,089**

(22) PCT Filed: **Jun. 23, 2004**

(86) PCT No.: **PCT/US04/20279**

§ 371 (c)(1),
(2), (4) Date: **Jan. 28, 2008**

(30) **Foreign Application Priority Data**

Jun. 23, 2003 (US) 10/603,138

Publication Classification

(51) **Int. Cl.**

C12Q 1/68 (2006.01)
C07H 21/04 (2006.01)

(52) **U.S. Cl. 435/6; 536/23.1**

(57) **ABSTRACT**

The invention provides methods, nucleic acids and kits for detecting colon cell proliferative disorders. The invention discloses genomic sequences the methylation patterns of which have utility for the improved detection of and differentiation between said class of disorders, thereby enabling the improved diagnosis and treatment of patients.

A

B

R H M R H M



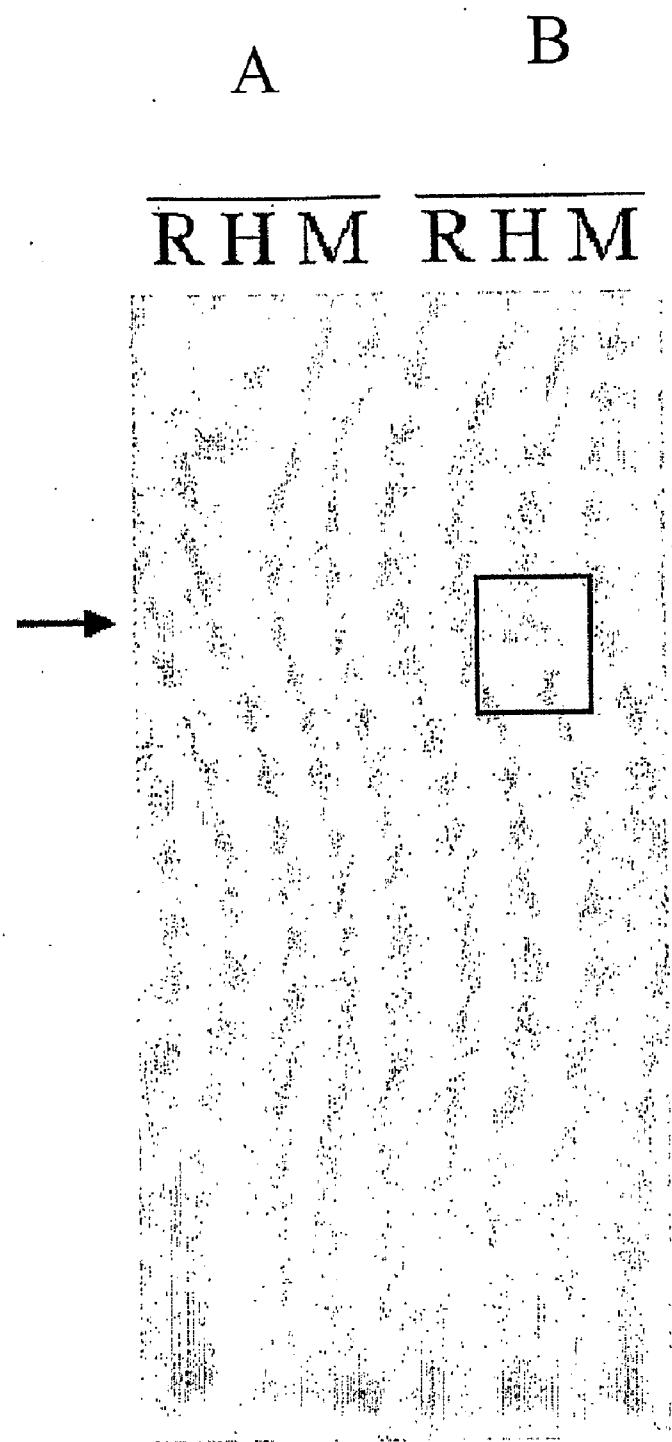


Figure 1

Figure 2

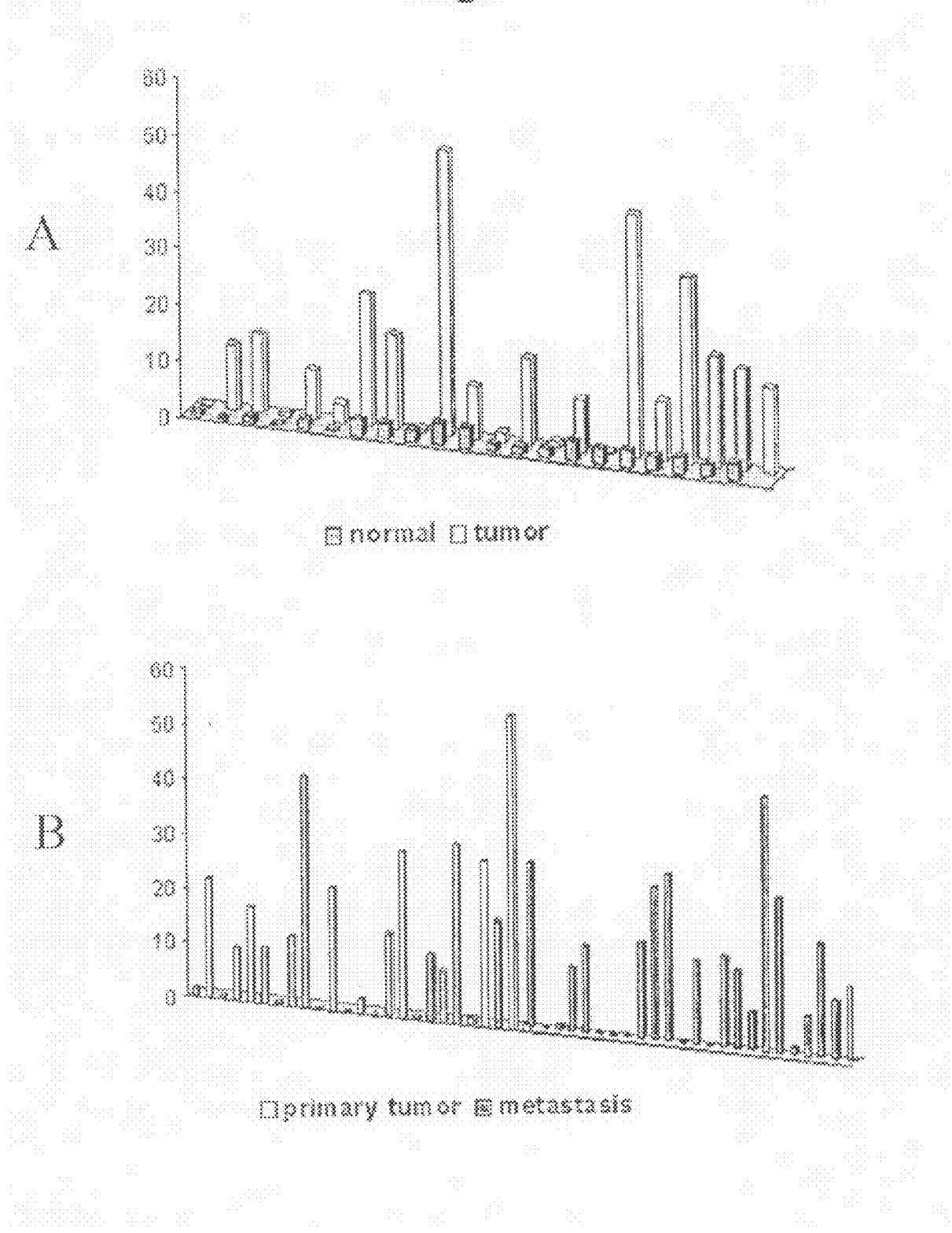
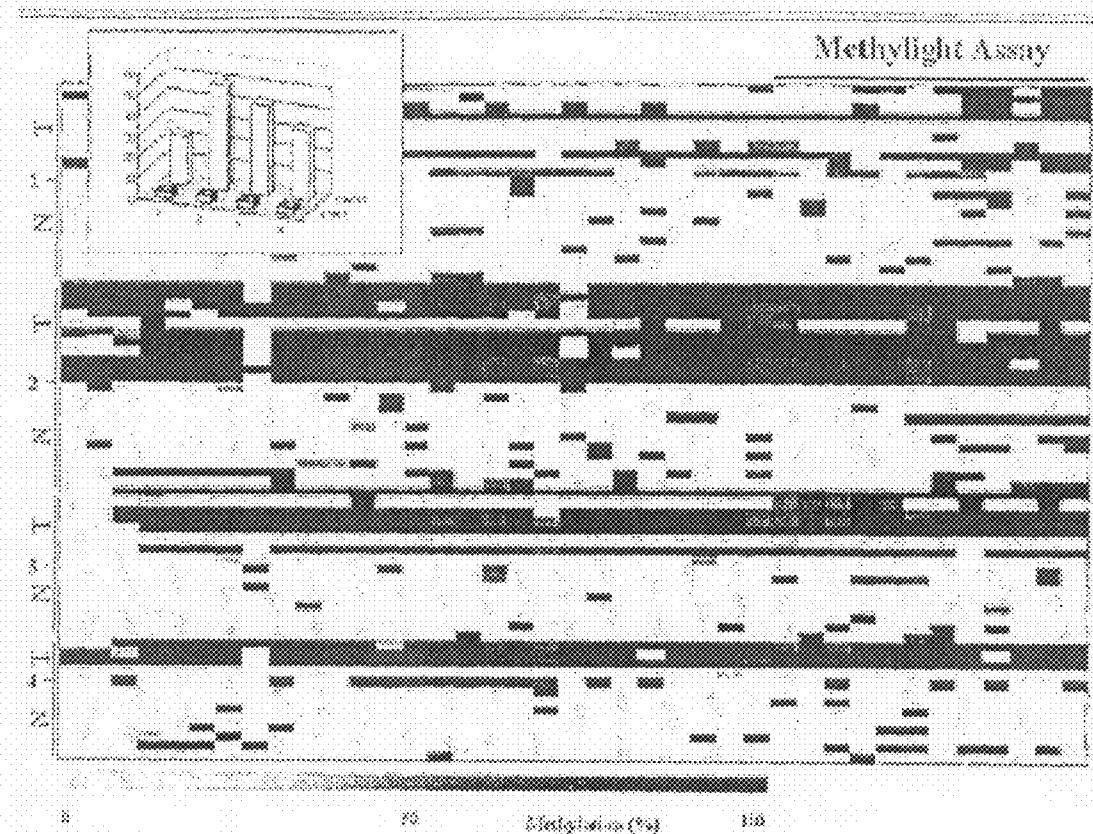


Figure 3



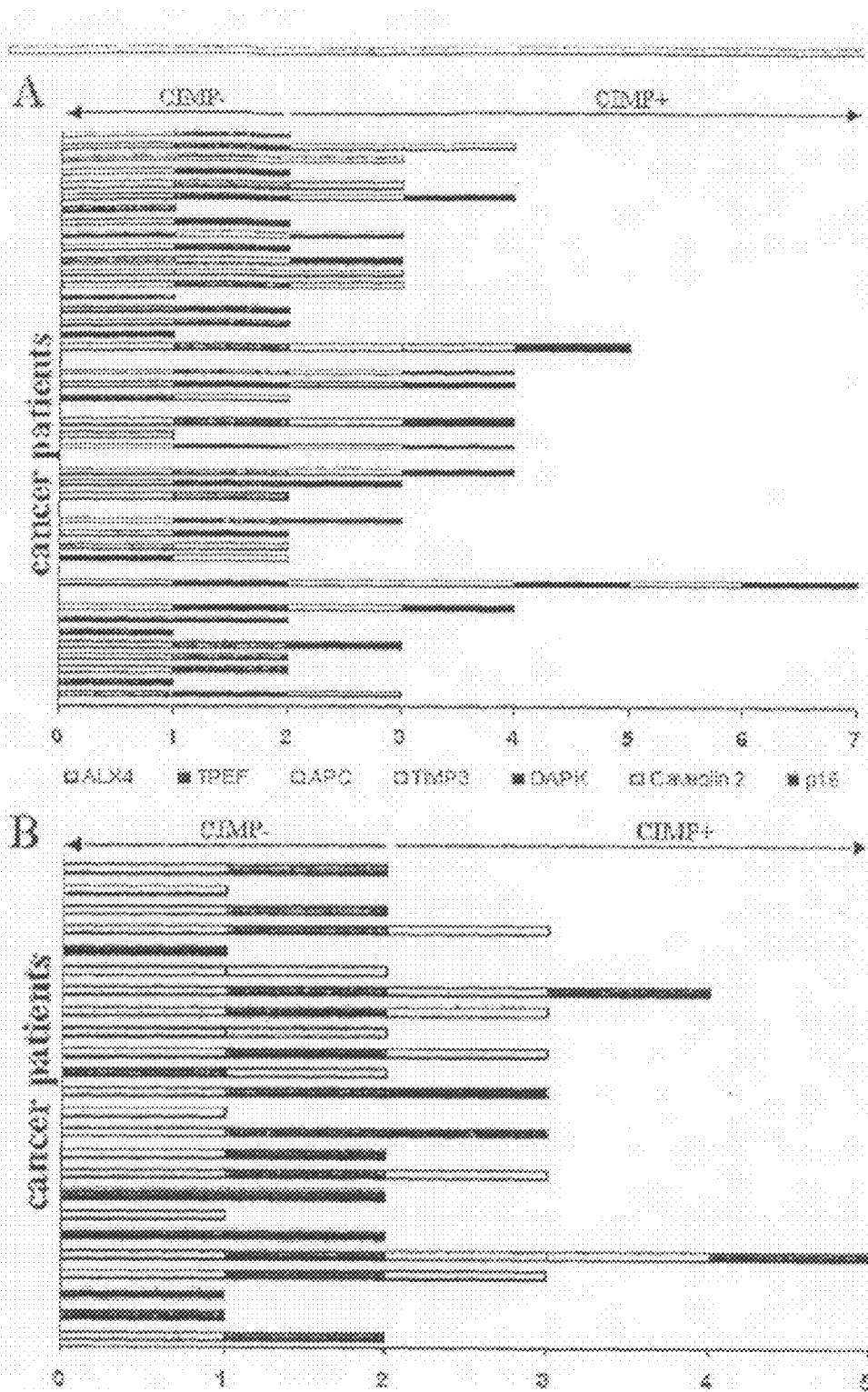
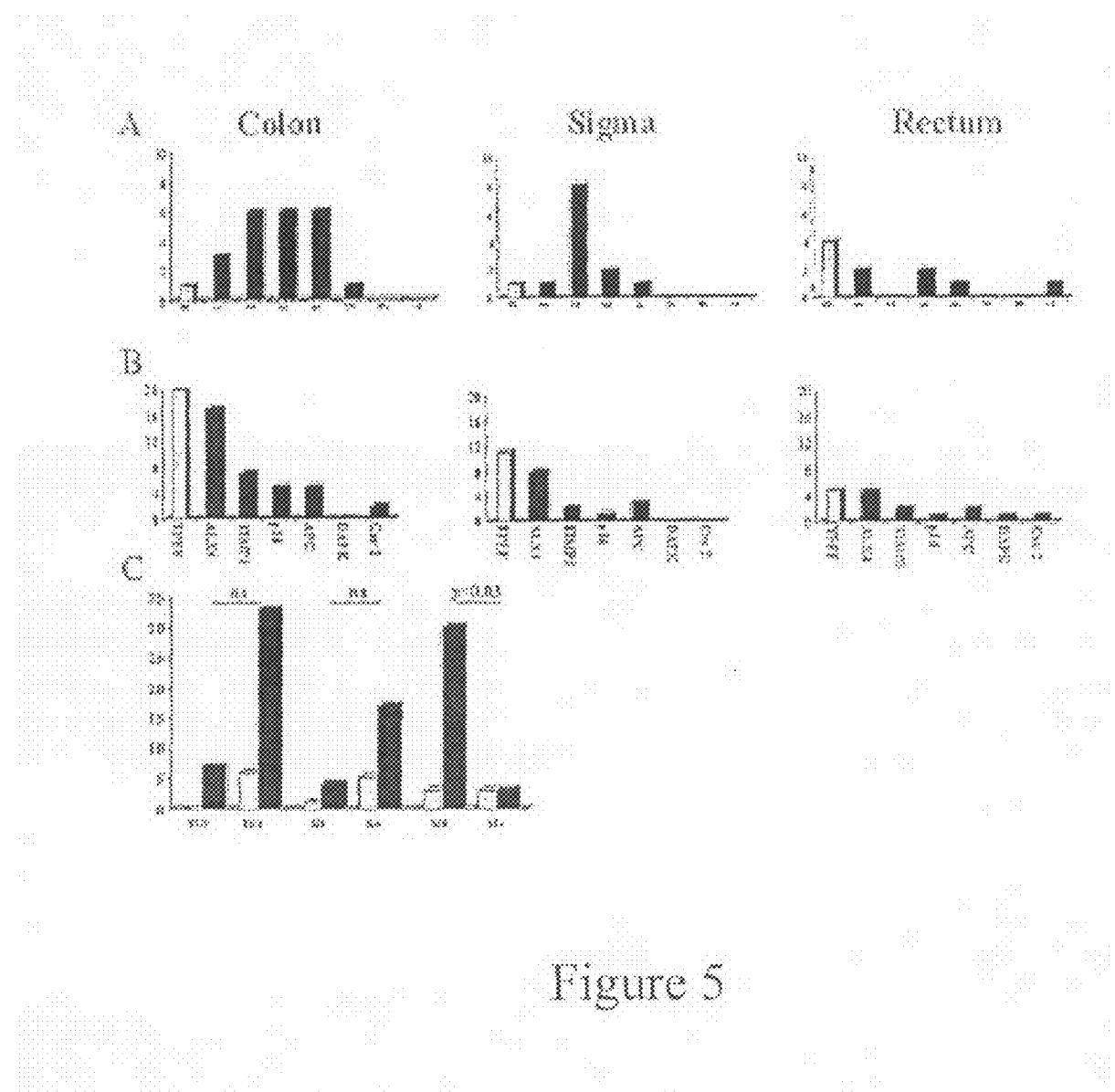


Figure 4



METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF COLON PROLIFERATIVE DISORDERS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. patent application Ser. No. 10/603,138, filed 23 June, 2003 and entitled METHODS AND NUCLEIC ACIDS FOR ANALYSES OF COLORECTAL CELL PROLIFERATIVE DISORDERS, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] Aspects of the present invention relate to cancer, and to the detection and progression of cancer. More particular aspects relate to nucleic acids and kits having diagnostic, prognostic and therapeutic utility for detecting and distinguishing colon cell proliferative disorders, based on methylation patterns of relevant genomic DNA sequences.

BACKGROUND

[0003] In the United States the annual incidence of colorectal cancer is approximately 150,000, with 56,600 individuals dying from colorectal cancer each year. The lifetime risk of colorectal cancer in the general population is about 5 to 6 percent. Despite intensive efforts in recent years in screening and early detection of colon cancer, until today most cases are diagnosed in an advanced stage with regional or distant metastasis. While the therapeutic options include surgery and adjuvant or palliative chemotherapy, most patients die from progression of their cancer within a few months. Identifying the molecular changes that underly the progression of colon cancer and the formation of metastasis may help to develop new diagnostic and therapeutic options that could improve the overall poor prognosis of these patients.

[0004] The current model of colorectal pathogenesis favours a stepwise progression of adenomas which includes the development of dysplasia and finally signs of invasive cancer. The molecular changes underlying this adenoma-carcinoma sequence include genetic and epigenetic alterations of tumor suppressor genes (APC, p 53, DCC), the activation of oncogenes (K-ras) and the inactivation of DNA mismatch repair genes¹. Recently, further molecular changes and genetic defects have been revealed. Thus, activation of the Wnt signalling pathway not only includes mutations of the APC gene, but may also result from beta-catenin mutations⁵. Furthermore, alterations in the TGF-beta signalling pathway together with its signal transducers SMAD4 and SMAD2 have been linked to the development of colon cancer.

[0005] Despite recent progress in the understanding of the pathogenesis of adenomas and carcinomas of the colon and their genetic and molecular changes, the genetic and epigenetic changes underlying the development of metastasis are less well understood. It is, however, generally well accepted that the process of invasion and proteolysis of the extracellular matrix, as well as infiltration of the vascular basement membrane involve adhesive proteins, such as members of the family of integrin receptors, the cadherins, the immunoglobulin superfamily, the laminin binding protein and the CD44 receptor. Apart from adhesion, the process of metastasis formation also includes the induction and regulation of angiogenesis (VEGF, bFGF), the induction of cell proliferation

(EGF, HGF, IGF) and the activation of proteolytic enzymes (MMPs, TIMPs, uPAR), as well as the inhibition of apoptosis (Bcl-2, Bcl-X). More recently other groups have compared the genetic and molecular changes in metastatic lesions to the changes found in primary colorectal cancers. Thus, Kleeff et al. reported the loss of DOC-2, a candidate tumor suppressor gene, both in primary and metastatic colorectal cancer. Furthermore, Zauber et al. reported that in their series of 42 colorectal cancers Ki-ras mutations in the primary cancers were identical in all of the 42 paired primary and synchronous metastatic lesions. Similarly loss of heterozygosity at the APC locus was identical for 39 paired carcinomas and synchronous metastasis. The authors concluded that for Ki-ras and APC genes the genetic changes in metastasis are identical to the primary colorectal cancer. However, other groups have found genetic and molecular changes in metastatic colon cancers, that are not present in the primary cancers. Thus, the development of LOH of chromosome 3p in colorectal metastasis has been reported. In addition, using comparative genomic hybridization several alterations were found in liver metastasis that were unique to metastatic lesions (-9q, -11q, and -17q)³⁸.

[0006] Apart from mutations aberrant methylation of CpG islands has been shown to lead to the transcriptional silencing of certain genes that have been previously linked to the pathogenesis of various cancers. CpG islands are short sequences which are rich in CpG dinucleotides and can usually be found in the 5' region of approximately 50% of all human genes. Methylation of the cytosines in these islands leads to the loss of gene expression and has been reported in the inactivation of the X chromosome and genomic imprinting. Recently several groups have also analysed the methylation of various genes in colorectal cancer and reported the transcriptional silencing by promoter methylation for p16INK4, p 14ARF, p 15INK4b, MGMT, hMLH1, GSTP1, DAPK, CDH1, TIMP-3 and APC among others. Thus apart from mutational inactivation of certain genes, the hypermethylation of these genes also contributes significantly to the pathogenesis of this disease.

[0007] In recent years several genes that are methylated in colon cancer have been identified by MS AP-PCR. This group of genes among others, includes TPEF/HPP1 which is frequently methylated in colon cancers and which was independently identified by two different groups using the MS AP-PCR method. See for example, Young J, Biden K G, Simms L A, Huggard P, Karamatic R, Eyre H J, Sutherland G R, Herath N, Barker M, Anderson G J, Fitzpatrick D R, Ramm G A, Jass J R, Leggett B A. HPP1: a transmembrane protein-encoding gene commonly methylated in colorectal polyps and cancers. Proc Natl Acad Sci USA 2001; 98:265-270.

[0008] ALX4 gene methylation was identified using differential methylation hybridization in a study by Yan et al. of genome-wide screening for CpG island hypermethylation in breast cancer samples. In their study ALX4 gene methylation was most prominent in poorly differentiated breast cancers (Yan P S, Perry M R, Laux D E, Asare A L, Caldwell C W, Huang T H. CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer. Clin Cancer Res 2000; 6:1432-1438). ALX4 is a putative transcription factor that belongs to the family of paired-class homeoproteins. This gene is part of a family of genes that includes the mammalian genes Alx 3, Cart-1, MHox, and S8 and exhibits similarity to the *Drosophila* gene aristless. It binds palindromic DNA sequences (5'-TAAT-3') as either homodimers or as heterodimers with other family members and strongly acti-

vates transcription from a promoter containing the homeodomain binding site, P2. ALX4 is expressed at several sites during development, including the craniofacial and limb-bud mesenchyme. Interestingly, ALX4 deficient mice exhibit body-wall defects, preaxial polydactyl), and a decreased size of the parietal plate of the skull, while mutations of the human homeobox gene ALX4 have been found in inherited defects of skull ossification. ALX4 is also expressed in various tissues whose development is dependent on epithelial-mesenchymal interactions and regulates mesenchymal-specific activities of LEF-1.

[0009] Multifactorial approach. Cancer diagnostics has traditionally relied upon the detection of single molecular markers (e.g. gene mutations, elevated PSA levels). Unfortunately, cancer is a disease state in which single markers have typically failed to detect or differentiate many forms of the disease. Thus, assays that recognize only a single marker have been shown to be of limited predictive value. A fundamental aspect of this invention is that methylation based cancer diagnostics and the screening, diagnosis, and therapeutic monitoring of such diseases will provide significant improvements over the state-of-the-art that uses single marker analyses by the use of a selection of multiple markers. The multiplexed analytical approach is particularly well suited for cancer diagnostics since cancer is not a simple disease, this multi-factorial “panel” approach is consistent with the heterogeneous nature of cancer, both cytologically and clinically.

[0010] Key to the successful implementation of a panel approach to methylation based diagnostic tests is the design and development of optimized panels of markers that can characterize and distinguish disease states. This patent application describes an efficient and unique panel of genes the methylation analysis of one or a combination of the members of the panel enabling the detection of cell proliferative disorders of the prostate with a particularly high sensitivity, specificity and/or predictive value.

[0011] Development of medical tests. Two key evaluative measures of any medical screening or diagnostic test are its sensitivity and specificity, which measure how well the test performs to accurately detect all affected individuals without exception, and without falsely including individuals who do not have the target disease (predictive value). Historically, many diagnostic tests have been criticized due to poor sensitivity and specificity.

[0012] A true positive (TP) result is where the test is positive and the condition is present. A false positive (FP) result is where the test is positive but the condition is not present. A true negative (TN) result is where the test is negative and the condition is not present. A false negative (FN) result is where the test is negative but the condition is present. In this context: Sensitivity=TP/(TP+FN); Specificity=TN/(FP+TN); and Predictive value=TP/(TP+FP).

[0013] Sensitivity is a measure of a test's ability to correctly detect the target disease in an individual being tested. A test having poor sensitivity produces a high rate of false negatives, i.e., individuals who have the disease but are falsely identified as being free of that particular disease. The potential danger of a false negative is that the diseased individual will remain undiagnosed and untreated for some period of time, during which the disease may progress to a later stage wherein treatments, if any, may be less effective. An example of a test that has low sensitivity is a protein-based blood test for HIV. This type of test exhibits poor sensitivity because it fails to detect the presence of the virus until the disease is well established

and the virus has invaded the bloodstream in substantial numbers. In contrast, an example of a test that has high sensitivity is viral-load detection using the polymerase chain reaction (PCR). High sensitivity is achieved because this type of test can detect very small quantities of the virus. High sensitivity is particularly important when the consequences of missing a diagnosis are high.

[0014] Specificity, on the other hand, is a measure of a test's ability to identify accurately patients who are free of the disease state. A test having poor specificity produces a high rate of false positives, i.e., individuals who are falsely identified as having the disease. A drawback of false positives is that they force patients to undergo unnecessary medical procedures treatments with their attendant risks, emotional and financial stresses, and which could have adverse effects on the patient's health. A feature of diseases which makes it difficult to develop diagnostic tests with high specificity is that disease mechanisms, particularly in cancer, often involve a plurality of genes and proteins. Additionally, certain proteins may be elevated for reasons unrelated to a disease state. An example of a test that has high specificity is a gene-based test that can detect a p53 mutation. Specificity is important when the cost or risk associated with further diagnostic procedures or further medical intervention are very high.

[0015] Bisulfite modification of DNA is an art-recognized tool used to assess CpG methylation status. 5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing, because 5-methylcytosine has the same base pairing behavior as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during, e.g., PCR amplification.

[0016] The most frequently used method for analyzing DNA for the presence of 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine whereby, upon subsequent alkaline hydrolysis, cytosine is converted to uracil which corresponds to thymine in its base pairing behavior. Significantly, however, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is converted in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using standard, art-recognized molecular biological techniques, for example, by amplification and hybridization, or by sequencing. All of these techniques are based on differential base pairing properties, which can now be fully exploited.

[0017] The prior art, in terms of sensitivity, is defined by a method comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of art-recognized methods for detecting 5-methylcytosine is provided by Rein, T., et al., *Nucleic Acids Res.*, 26:2255, 1998.

[0018] The bisulfite technique, barring few exceptions (e.g., Zeschnigk M, et al., *Eur J Hum Genet.* 5:94-98, 1997), is currently only used in research. In all instances, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment, and either completely sequenced (Olek & Walter, *Nat. Genet.* 1997 17:275-6, 1997), subjected to one or more primer extension reactions (Gonzalgo & Jones, *Nucleic Acids Res.*, 25:2529-31, 1997; WO 95/00669; U.S. Pat. No. 6,251,594) to analyze individual cytosine positions, or treated by enzymatic digestion (Xiong & Laird, *Nucleic Acids Res.*, 25:2532-4, 1997). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark, *Bioessays*, 16:431-6, 1994; Zeschnigk M, et al., *Hum Mol. Genet.*, 6:387-95, 1997; Feil R, et al., *Nucleic Acids Res.*, 22:695-, 1994; Martin V, et al., *Gene*, 157:261-4, 1995; WO 9746705 and WO 9515373).

[0019] Bisulfite Methylation Assay Procedures. Various methylation assay procedures are known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (e.g., CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

[0020] For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, e.g., the method described by Sadri & Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

[0021] COBRA. COBRATM analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples.

[0022] Other assays used in the art include "MethylLightTM" (a fluorescence-based real-time PCR technique) (Eads et al., *Cancer Res.* 59:2302-2306, 1999), Ms-SNu-PETTTM (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP";

Hernan et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; U.S. Pat. No. 5,786,146), and methylated CpG island amplification ("MCA"; Toyota et al., *Cancer Res.* 59:2307-12, 1999). These may be used alone or in combination with other of these methods.

[0023] MethylLight. The MethylLightTM assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan[®]) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999). Briefly, the MethylLightTM process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

[0024] The MethylLightTM assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

[0025] The MethylLightTM process can be used with a "Taq-Man[®]" probe in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan[®] probes; e.g., with either biased primers and TaqMan[®] probe, or unbiased primers and TaqMan[®] probe. The TaqMan[®] probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10° C. higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan[®] probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan[®] probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan[®] probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

[0026] Alternatively the MethylLightTM process can be used with 'LightCycler' probes. A LightCyclerTM probe is a pair of single-stranded fluorescent-labeled oligonucleotides. The first oligonucleotide probe is labeled at its 3' end with a donor fluorophore dye and the second is labeled at its 5' end with an acceptor fluorophore dye. The free 3' hydroxyl group of the second probe is blocked with a phosphate group to prevent polymerase mediated extension.

[0027] During the annealing step of real-time quantitative PCR, the PCR primers and the LightCyclerTM probes hybrid-

ize to their specific target regions causing the donor dye to come into close proximity to the acceptor dye. When the donor dye is excited by light, energy is transferred by Fluorescence Resonance Energy Transfer (FRET) from the donor to the acceptor dye. The energy transfer causes the acceptor dye to emit fluorescence wherein the increase of measured fluorescence signal is directly proportional to the amount of target DNA.

[0028] Typical reagents (e.g., as might be found in a typical MethylLight™-based kit) for MethylLight™ analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); Taq-Man® and/or LightCycler™ probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

[0029] Ms-SNuPE. The Ms-SNuPETM technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site(s) of interest. Small amounts of DNA can be analyzed (e.g., microdissected pathology sections), and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

[0030] Typical reagents (e.g., as might be found in a typical Ms-SNuPETM-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPETM primers for specific gene; reaction buffer (for the Ms-SNuPETM reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

[0031] MSP. MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herran et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; U.S. Pat. No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. This technique has been described in U.S. Pat. No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methy-

lated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

[0032] Pronounced need in the art. Therefore, in view of the incidence of colon cancer there is a substantial need in the art for the development of molecular markers that could be used for the early detection of colorectal cell proliferative disorders, in particular colon cancer. Additionally, there is a pronounced need in the art for the development of molecular markers that could be used to provide sensitive, accurate and non-invasive methods (as opposed to, e.g., biopsy) for the diagnosis, prognosis and treatment of colon cell proliferative disorders.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] FIG. 1 shows the results of the MS-APPCR analysis of pooled DNA from normal and adenoma tissues. A hypermethylated gene (ALX4) was detected in the adenoma DNA (box, arrow). R, Rsal; H, Rsal/HpaII; M, Rsal/MspI.

[0034] FIGS. 2A and 2B show the degree of methylation of the ALX4 gene as assessed by MethylLight™ assay as outlined in examples. The X-axis shows the percentage methylated reference (PMR). Figure A shows ALX4 gene methylation in normal colon mucosa and matched colon cancer; Figure B shows ALX4 gene methylation in primary colorectal cancers and metastasis.

[0035] FIG. 3 shows bisulfite sequencing. Four cases of cancers with a high degree of methylation and their matched normal non-malignant colon mucosa tissues were selected (inset). ALX4 gene methylation was confirmed by sequencing of bisulfite treated genomic DNA of samples 1-4. N, normal mucosa; T, tumor, numbers 1-4 correspond to cases 1-4 in the Methylight assay (inset); Methylight assay, indicates the CpG sites that were covered by both the Methylight assay and sequencing of the respective DNA fragments.

[0036] FIGS. 4A and 4B show the number of methylated genes per primary colorectal cancer (A) and metastasis (B). CIMP+ status was defined as 2 or more methylated genes per patient.

[0037] FIGS. 5A, 5B and 5C show an analysis of methylation with regard to clinicopathological features of colorectal cancers. A. In cancers of the colon and the sigma methylation was significantly more frequent than in cancers of the rectum only. y axis: number of patients; x axis: number of methylated genes B. TPEF methylation was significantly more frequent in cancers of the colon and sigma. Number of patients analysed: colon cancer: n=23, sigma cancer: n=13, rectal cancer: n=10. One patient with recurrent cancer was not included in this analysis. y axis: number of patients; x axis: number of methylated genes. C. Cancers without distant metastasis were frequently methylated, grey box: no methylation; black box: methylation of 1 to 7 genes.

DETAILED DESCRIPTION OF THE INVENTION

[0038] For the purposes of the following invention the 'sensitivity' and 'specificity' refer to values calculated by reference to a sample set according to that described in the EXAMPLES disclosed herein.

Definitions

[0039] The term "Observed/Expected Ratio" ("O/E Ratio") refers to the frequency of CpG dinucleotides within a particu-

lar DNA sequence, and corresponds to the [number of CpG sites/(number of C bases×number of G bases)]×band length for each fragment.

[0040] The term “CpG island” refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an “Observed/Expected Ratio”>0.6, and (2) having a “GC Content”>0.5. CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

[0041] The term “methylation state” or “methylation status” refers to the presence or absence of 5-methylcytosine (“5-mCyt”) at one or a plurality of CpG dinucleotides within a DNA sequence. Methylation states at one or more particular palindromic CpG methylation sites (each having two CpG CpG dinucleotide sequences) within a DNA sequence include “unmethylated,” “fully-methylated” and “hemi-methylated.”

[0042] The term “hemi-methylation” or “hemimethylation” refers to the methylation state of a palindromic CpG methylation site, where only a single cytosine in one of the two CpG dinucleotide sequences of the palindromic CpG methylation site is methylated (e.g., 5'-NC^MGN-3' (top strand); 3'-NGCN-5' (bottom strand)).

[0043] The term “hypermethylation” refers to the average methylation state corresponding to an increased presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

[0044] The term “hypomethylation” refers to the average methylation state corresponding to a decreased presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

[0045] The term “microarray” refers broadly to both “DNA microarrays,” and ‘DNA chip(s),’ as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon. “Genetic parameters” are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

[0046] “Epigenetic parameters” are, in particular, cytosine methylations. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analyzed using the described method but which, in turn, correlate with the DNA methylation.

[0047] The term “bisulfite reagent” refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

[0048] The term “Methylation assay” refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

[0049] The term “MS AP-PCR” (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Goncalgo et al., *Cancer Research* 57:594-599, 1997.

[0050] The term “MethyLight®” refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

[0051] The term “HeavyMethy™” assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl MethylLight® assay, which is a variation of the MethylLight® assay, wherein the MethylLight® assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers.

[0052] The term “Ms-SNuPE” (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

[0053] The term “MSP” (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by U.S. Pat. No. 5,786,146.

[0054] The term “COBRA” (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

[0055] The term “MCA” (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al., *Cancer Res.* 59:2307-12, 1999, and in WO 00/26401A1.

[0056] The term “hybridization” is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

[0057] “Stringent hybridization conditions,” as defined herein, involve hybridizing at 68° C. in 5×SSC/5×Denhardt’s solution/1.0% SDS, and washing in 0.2×SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60° C. in 2.5×SSC buffer, followed by several washing steps at 37° C. in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3×SSC at 42° C., or the art-recognized equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N. Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N. Y.) at Unit 2.10.

[0058] The term ‘primary’ when used in reference to cancer or other cell proliferative disorder shall be taken to mean the first to develop.

[0059] The term ‘metastasis’ as used herein shall be taken to mean the transfer of a disease-producing agent (such as bacteria, cancer or other cell proliferative disorder cells) from an original site of disease to another part of the body with development of a similar lesion in the new location.

Overview:

[0060] Despite intensive efforts to improve screening and early detection of colon cell proliferative disorders, most cases are diagnosed in an advanced stage with regional or distant metastasis which are associated with poor survival. The herein described invention discloses epigenetic markers that have novel utility for the analysis of colon cell proliferative disorders, combined with sensitive assay methods for the improved detection of said disorders. The invention presents

improvements over the state of the art in that it provides a means for the detection of colon cell proliferative disorders by analysis of a gene panel, with a high sensitivity and specificity. The invention presents further improvements in that the ‘gene panel’ consists of at least one of seven genes and/or their regulatory sequences, thereby enabling a highly sensitive and specific but time and cost effective analysis. The invention further discloses particularly preferred combinations of said seven genes.

[0061] In one aspect, the present invention provides for the improved detection of colorectal carcinomas by determination of the methylation status of CpG dinucleotide sequences of the gene ALX4 and/or its regulatory sequences. In a further aspect the invention provides a further preferred means for the detection of colorectal carcinomas by determination of the methylation status of CpG dinucleotide sequences of the gene ALX4 and at least one gene selected from the group consisting TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences.

[0062] In one aspect, the present invention provides for the use of the bisulfite technique, in combination with one or more methylation assays, for determination of the methylation status of CpG dinucleotide sequences of at least two genes taken from the group consisting ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences. It is particularly preferred that one of the genes is ALX4. According to the present invention, determination of the methylation status of CpG dinucleotide sequences within at least two members of said group of genes has diagnostic and prognostic utility. It is a further aspect of the invention that the analysed genomic sequences of the group consisting the group consisting ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences are selected from the sequence listing (see Table 2).

[0063] In a further aspect the present invention provides a selection of genes consisting ALX4, TPEF and p16. Two or more of these genes are analysed in the form of a ‘gene panel’. It is particularly preferred that one of the genes is ALX4. It is a further aspect of the invention that the analysed genomic sequences are selected from the group consisting of ALX4, TPEF and p16 and/or their regulatory sequences are selected from the sequence listing (see Table 2).

[0064] Particular embodiments of the present invention provide a novel application of the analysis of methylation levels and/or patterns within said sequences that enables a precise detection, classification, treatment and overall prognosis of colon cell proliferative disorders. Early detection of colon cell proliferative disorders is directly linked with disease prognosis, and the disclosed method thereby enables the physician and patient to make better and more informed treatment decisions.

More Particularly:

[0065] The present invention provides improved means for the detection of colorectal cell proliferative disorders. This aim is achieved by the analysis of the CpG methylation status of genes selected from the group consisting ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences.

[0066] In a further aspect, the present invention achieves said goal by analysis of the methylation status of at least one CpG position of the gene ALX4 and/or its regulatory sequences. In a further aspect the aim of the invention is

achieved by the methylation analysis of said gene, ALX4 and/or its regulatory sequences and one or more genes selected from the group consisting TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences.

[0067] The present invention is further based upon the analysis of methylation levels within two or more genes taken from the group consisting of ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences.

[0068] Accordingly, the invention also disclose the genomic sequences of said genes in SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NO:45 TO SEQ ID NO:47, according to table 2.

[0069] Additional embodiments provide modified variants of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation patterns within SEQ ID NO:1 to SEQ ID NO:4 and SEQ ID NO:45 to SEQ ID NO:47.

[0070] According to the present invention hypermethylation of the genes ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences is correlated with varying degrees of probability to the presence of colon cell proliferative disorders, and or metastases thereof. The present invention discloses the analysis of methylation within said genes and/or their regulatory sequences in the form of a panel enabling the improved detection, classification, treatment and overall prognosis of colon cell proliferative disorders. Aberrant methylation of the genes TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 have to date been associated with the development of colorectal cell proliferative disorders. The present invention provides specific combinations of these genes which were determined to be particularly useful for the detection of colorectal cell proliferative disorders as measured by sensitivity and specificity of detection. Furthermore the invention provides CpG methylation analysis of the gene ALX4, with specific and novel utility for the detection of colorectal cell proliferative disorders. Methylation analysis of this gene is herein shown to have the surprising effect of being a highly sensitive and specific colorectal cancer detection marker. Furthermore the sensitivity and specificity of this detection is improved by a combined analysis of the gene ALX4 and one or more genes selected from the group consisting of TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3.

[0071] Wherein the object of the analysis is the detection of colon cell proliferative disorders it is particularly preferred that the methylation of two or more genes selected from the group consisting of ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences are analysed. It is particularly preferred that said genes are selected from the group consisting of ALX4, TPEF and p16 and/or their regulatory sequences are analysed. It is further preferred that the methylation of all of the genes of the group consisting ALX4, TPEF and p16 and/or their regulatory sequences are analysed. In an alternative embodiment the methylation of the gene ALX4 only is analysed. In a further preferred alternative embodiment the CpG methylation status of the gene ALX4 and/or its regulatory sequences and one or more genes selected from the group consisting of TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences are analysed.

[0072] An objective of the invention comprises analysis of the methylation state of two or more CpG dinucleotides within at least two of the genomic sequences selected from the group consisting of SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 45 to SEQ ID NO: 47 and sequences complementary thereto.

[0073] It is preferred that the methylation of two or more sequences selected from the group consisting SEQ ID NOS: 2, 3 and 4 are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the sequences of the group consisting SEQ ID NOS: 2, 3 and 4 are analysed.

[0074] In an alternative embodiment the methylation status of at least one CpG position of SEQ ID NO: 2 only is analysed. In a further preferred alternative embodiment the CpG methylation status of SEQ ID NO: 2 and one or more sequences selected from the group consisting SEQ ID NOS: 1, 3, 4 and SEQ ID NOS: 45 to SEQ ID NO: 47 and sequences complementary thereto are analysed. In a further preferred alternative embodiment the CpG methylation status of SEQ ID NO: 2 and one or more sequences selected from the group consisting SEQ ID NOS: 1, 3 and 4 and sequences complementary thereto are analysed.

[0075] The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 45 to SEQ ID NO: 47, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more, consecutive or random methylated CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof. In a preferred embodiment of the invention, the objective comprises analysis of at least two modified nucleic acid comprising a sequence of at least 16 contiguous nucleotide bases in length of a sequence selected from the group consisting of SEQ ID NO: 5 TO SEQ ID NO: 20 & SEQ ID NO: 48 TO SEQ ID NO: 59, wherein said sequence comprises at least one CpG, TpA or CpA dinucleotide and sequences complementary thereto. The sequences of SEQ ID NOS: 7 to SEQ ID NO: 12 and SEQ ID NOS: 15 to SEQ ID NO: 20 provide modified versions of the nucleic acid according to SEQ ID NOS: 2 to SEQ ID NO: 4, wherein the modification of each genomic sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from said genomic sequence as follows. For each sense strand genomic DNA, e.g., SEQ ID NO: 1, four converted versions are disclosed. A first version wherein "C" is converted to "T," but "CpG" remains "CpG" (i.e., corresponds to case where, for the genomic sequence, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); a second version discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C" is converted to "T," but "CpG" remains "CpG" (i.e., corresponds to case where, for all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted). The 'upmethylated' converted sequences of SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 45 to SEQ ID NO: 47 correspond to SEQ ID NOS: 5 to SEQ ID NO: 12 and SEQ ID NOS: 48 to SEQ ID NO: 53 (see TABLE 2). A third chemically converted version of each genomic sequences is provided, wherein "C" is converted to "T" for all "C" residues, including those of "CpG" dinucle-

otide sequences (i.e., corresponds to case where, for the genomic sequences, all "C" residues of CpG dinucleotide sequences are unmethylated); a final chemically converted version of each sequence, discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C" is converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the complement (antisense strand) of each genomic sequence, all "C" residues of CpG dinucleotide sequences are unmethylated). The 'downmethylated' converted sequences of SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 45 to SEQ ID NO: 47 correspond to SEQ ID NOS: 13 to SEQ ID NO: 20 and SEQ ID NOS: 54 to SEQ ID NO: 59.

[0076] Particularly useful for the detection of colon cell proliferative disorders, and heretofore undisclosed are the non-naturally occurring sequences according to SEQ ID NOS: 7, 8, 15 and 16, which correspond to methylation-specific converted sequences of part of the gene ALX4 (SEQ ID NO: 2).

[0077] In an alternative preferred embodiment, such analysis comprises the use of an oligonucleotide or oligomer for detecting the cytosine methylation state within genomic or pretreated (chemically modified) DNA, according to SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 48 to SEQ ID NO: 59. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a pretreated nucleic acid sequence according to SEQ ID NOS: 5 to SEQ ID NO: 20 & SEQ ID NOS: 48 to SEQ ID NO: 59 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 45 to SEQ ID NO: 47 and/or sequences complementary thereto.

[0078] Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NOS: 1 to SEQ ID NO: 20, or to the complements thereof. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

[0079] Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NOS: 1 to SEQ ID NO: 20, or to the complements thereof.

[0080] Hybridizing nucleic acids of the type described herein can be used, for example, as a primer (e.g., a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or T_m , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

[0081] For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NOS: 1 to SEQ ID NO: 4 and SEQ ID NOS: 45 to SEQ ID NO: 47 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at

which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a 1° C. decrease in the Tm, the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having >95% identity with the probe are sought, the final wash temperature is decreased by 5° C.). In practice, the change in Tm can be between 0.5° C. and 1.5° C. per 1% mismatch.

[0082] Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NO: 1, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X, where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

[0083] n to (n+(X-1));

[0084] where n=1, 2, 3, . . . (Y-(X-1));

[0085] where Y equals the length (nucleotides or base pairs);

[0086] where X equals the common length (in nucleotides) of each oligonucleotide in the set (e.g., X=20 for a set of consecutively overlapping 20-mers); and where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to Y-(X-1).

[0087] Preferably, the set is limited to those oligomers that comprise at least one CpQ TpG or CpA dinucleotide.

[0088] Examples of inventive 20-mer oligonucleotides within a sequence of length 2470 base pairs include the following set of 2470 oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions 1-20, 2-21, 3-22, 4-23, 5-24 . . . 2451-2470.

[0089] Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

[0090] The present invention encompasses, for each of SEQ ID NO: 1 to SEQ ID NO: 20 and SEQ ID NO: 45 to SEQ ID NO: 59 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, e.g., X=9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

[0091] The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO 47. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO:1 to SEQ ID NO:20 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

[0092] Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

[0093] The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophors, lipids such as cho-

lesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, U.S. Pat. Nos. 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly p referred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a chromophore, fluorophor, peptide, hybridization-triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

[0094] The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

[0095] The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in 'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequence SEQ ID NO: 2 to SEQ ID NO: 4 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the pretreated nucleic acids according to SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20 and sequences complementary thereto. In a further preferred embodiment the set comprises contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequence SEQ ID NO:2 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the pretreated nucleic acids according to SEQ ID NOS: 7, 8, 15 and 16.

[0096] However, it is anticipated that for economic or other factors it may be preferable to analyze a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

[0097] Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state in pretreated genomic DNA of at least two genes selected from ALX4, TPEF and p16 (SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20), or in genomic DNA (SEQ ID NOS:2 to SEQ ID NO:4 and sequences complementary thereto). In a further preferred embodiment the set comprises at least two oligonucleotides for the analysis of CpG positions within one or more of SEQ ID NOS:7, 8, 15 and 16.

[0098] These probes enable diagnosis, classification and/or therapy of genetic and epigenetic parameters of colon cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20), or in genomic DNA (SEQ ID NOS:2 to SEQ ID NO:4 and sequences complementary thereto).

[0099] In preferred embodiments, at least one, and more preferably all members of a set of oligonucleotides is bound to a solid phase.

[0100] In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of

one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:49 and sequences complementary thereto, or segments thereof.

[0101] It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (i.e., an arrangement of different oligonucleotides and/or PNA-oligomers bound to a solid phase). Such an array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized, for example, in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid-phase surface may be composed of silicon, glass, polystyrene, aluminum, steel, iron, copper, nickel, silver, or gold. Nitrocellulose as well as plastics such as nylon, which can exist in the form of pellets or also as resin matrices, may also be used. An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of *Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999, and from the literature cited therein)*. Fluorescently labeled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridized probes may be carried out, for example, via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

[0102] It is particularly preferred that the oligomers according to the invention are utilised for the detection of colorectal carcinoma.

[0103] The present invention further provides a method for ascertaining genetic and/or epigenetic parameters of the genes ALX4, TPEF, p16, APC, TIMP3, Dapk and Caveolin 2 and/or their regulatory sequences within a subject by analyzing cytosine methylation and single nucleotide polymorphisms.

[0104] It is preferred that the methylation of two or more genes selected from the group consisting of the genes ALX4, TPEF, p16, APC, TIMP3, Dapk and Caveolin 2 and/or their regulatory sequences are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the genes of the group consisting of ALX4, TPEF and p16 and/or their regulatory sequences are analysed.

[0105] In a further embodiment of the method it is preferred that only the methylation status of the gene ALX4 and/or its regulatory sequences is analysed. In a further preferred embodiment the methylation status of the gene ALX4 and/or its regulatory sequences and one or more of the group consisting of TPEF, p16, APC, TIMP3, Dapk and Caveolin 2 and/or their regulatory sequences are analysed. In a further preferred embodiment the methylation status of the gene ALX4 and/or its regulatory sequences and one or both of the group consisting of TPEF and p16 and/or their regulatory sequences are analysed.

[0106] Accordingly, it is preferred that the methylation of two or more genetic sequences selected from the group consisting of SEQ ID NOS: 2,3 and 4 and/or their regulatory sequences are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the genes of the group consisting of SEQ ID NOS: 2,3 & 4 and/or their regulatory sequences are analysed.

[0107] In a further embodiment of the method it is preferred that only the methylation status of SEQ ID 2 is analysed. In a further preferred embodiment the methylation status of SEQ ID 2 and one or more of the group consisting of SEQ ID NOS: 3,

4, 45-47 are analysed. In a further preferred embodiment the methylation status of SEQ ID NO:2 and one or both of SEQ ID NOS:3 and 4 are analysed.

[0108] Said method comprising contacting a nucleic acid comprising one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 in a biological sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

[0109] It is preferred that the methylation of two or more genetic sequences select analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the genes of the group consisting SEQ ID NOS:2, 3 and 4 and/or their regulatory sequences are analysed.

[0110] In a further embodiment of the method it is preferred that only the methylation status of SEQ ID 2 is analysed. In a further preferred embodiment the methylation status of SEQ ID NO:2 and one or more of the group consisting SEQ ID NOS:3, 4 45-47 are analysed. In a further preferred embodiment the methylation status of SEQ ID NO:2 and one or both of SEQ ID NOS:3 and 4 are analysed.

[0111] Said method comprising contacting a nucleic acid comprising the appropriate gene(s) and/or one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 in a biological sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

[0112] Preferably, said method comprises the following steps: In the first step, a sample of the tissue to be analysed is obtained. The source may be any suitable source, such as cell lines, histological slides, biopsies, tissue embedded in paraffin, bodily fluids, stool, blood and all possible combinations thereof. Genomic DNA is then isolated from said biological sample, this may be by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants, e.g., by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA.

[0113] Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

[0114] In the second step of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as 'pretreatment' herein.

[0115] The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior.

[0116] In the third step of the method, fragments of the pretreated DNA are amplified, using sets of primer oligo-

nucleotides according to the present invention, and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20 and sequences complementary thereto.

[0117] It is preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NOS:7-12 and 15-20. In this embodiment of the invention it is particularly preferred that said set consists of at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of SEQ ID NOS:5-20 and 48-59.

[0118] It is also preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NOS:7-12 and 15-20.

[0119] It is also preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of at least one of SEQ ID NOS:7, 8, 15 or 16. It is also preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of at least one of SEQ ID NOS:7, 8, 15 or 16, and a further pair of oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of at least one of SEQ ID NOS:9-12 and 17-20.

[0120] In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in U.S. Pat. No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID

NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

[0121] It is preferred that said set of MSP primer oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7-12, 15-20 and 49-59.

[0122] In this embodiment of the invention it is particularly preferred that said set consists of at least two MSP primer oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of SEQ ID NOS:7-12 and 15-20.

[0123] In a further embodiment of the invention it is particularly preferred that said set consists of at least two MSP primer oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of at least one of the base sequences of SEQ ID NOS:7, 8, 15 and 16. In a further preferred embodiment of this method it is preferred that said set consists of at least two MSP primer oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of at least one of the base sequences of SEQ ID NOS:7, 8, 15 and 16 and one or more sequences taken from the group consisting SEQ ID NOS:9-12, 17-20 and 48-59.

[0124] A further preferred embodiment of the method comprises the use of blocker oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'CpA' or 'TpA' at the position in question, as opposed to a 'CpG' if the suppression of amplification of methylated nucleic acids is desired.

[0125] For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivitized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

[0126] Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-terminii thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and

primer-binding sites overlap, thereby precluding binding of the primer (e.g., with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker—a process that normally results in degradation of the hybridized blocker oligonucleotide.

[0127] A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

[0128] Preferably, therefore, the base sequence of said blocking oligonucleotides is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59 and sequences complementary thereto, wherein the base sequence of said oligonucleotides comprises at least one CpG TpG or CpA dinucleotide.

[0129] It is preferred that said set of blocking oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7-12 and 15-20. In this embodiment of the invention it is particularly preferred that said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7-12 and 15-20.

[0130] In a further embodiment it is preferred that said set of blocking oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7, 8, 15 & 16. In this embodiment of the invention it is particularly preferred that said set further consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:9-12, 17-20 & 48-59. In this embodiment it is further preferred that said base sequences are selected from SEQ ID NOS:9-12 and 17-20.

[0131] The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplificates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, e.g., matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

[0132] Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas & Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the

vapour phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut & Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionately with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut & Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

[0133] In the fourth step of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

[0134] In embodiments where the amplificates were obtained by means of MSP amplification, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primer, according to the base sequences of said primer.

[0135] Amplificates obtained by means of both standard and methylation specific PCR may be further analyzed by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

[0136] In one embodiment of the method, the amplificates synthesised in step three are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligonucleotides; in the process, the amplificates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the present Sequence Listing; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

[0137] In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, and the equivalent positions within SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO: 20. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplifies are then removed. The hybridized amplifies are then detected. In this context, it is preferred that labels attached to the amplifies are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

[0138] In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

[0139] A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; also see U.S. Pat. No. 6,331,393) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan™ PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Applied Biosystems, Foster City, Calif.). The TaqMan™ PCR reaction employs the use of a nonextendible interrogating oligonucleotide, called a TaqMan™ probe, which, in preferred embodiments, is designed to hybridize to a GpC-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (e.g., phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in U.S. Pat. No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethylLight™ assay. Variations on the TaqMan™ detection methodology that are also suitable for use with the described invention include the use of dual-probe technology (Lightcycler™) or fluorescent amplification primers (Sunrise™ technology). Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

[0140] A further suitable method for the use of probe oligonucleotides for the assessment of methylation by analysis of bisulfite treated nucleic acids In a further preferred embodiment of the method, the fifth step of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

[0141] In yet a further embodiment of the method, the fifth step of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the third step of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

Best Mode

[0142] In the most preferred embodiment of the method the nucleic acids according to SEQ ID NO: 1 to SEQ ID NO: 4

and SEQ ID NO: 45 to SEQ ID NO: 47 are isolated and treated according to the first three steps of the method outlined above, namely:

[0143] a. obtaining, from a subject, a biological sample having subject genomic DNA;

[0144] b. extracting or otherwise isolating the genomic DNA;

[0145] c. treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;

and wherein the subsequent amplification of d) is carried out in a methylation specific manner, namely by use of methylation specific primers or blocking oligonucleotides, and further wherein the detection of the amplificates is carried out by means of a real-time detection probes, as described above.

[0146] It is particularly preferred that the methylation of one or more sequences selected from the group consisting SEQ ID NOS: 2,3 and 4 are analysed. In this embodiment of the invention it is further preferred that the methylation of all of the sequences of the group consisting SEQ ID NOS: 2,3 and 4 are analysed.

[0147] In a further embodiment it is preferred that only the methylation status of SEQ ID NO:2 is analysed. In a further embodiment it is preferred that the methylation status of SEQ ID NO:2 and one or more sequences selected from the group consisting SEQ ID NOS:3, 4, 45-47 are analysed, even more preferably said group consists SEQ ID NOS:3 and 4 only.

[0148] Wherein the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

[0149] It is preferred that said set of MSP primer oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7-12 and 15-20.

[0150] In a further embodiment it is preferred that the said set MSP primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 & 16.

[0151] In a further preferred variant of said embodiment said set consists of at least MSP primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 and 16 and at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:9-12 and 17-20.

[0152] Step e) of the method, namely the detection of the specific amplificates indicative of the methylation status of one or more CpG positions according to SEQ ID NOS:1 to

SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 is carried out by means of real-time detection methods as described above.

[0153] In an alternative most preferred embodiment of the method the subsequent amplification of d) is carried out in the presence of blocking oligonucleotides, as described above. Said blocking oligonucleotides comprising a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS: 5-20 and 48-59 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide.

[0154] It is preferred that said set of blocking oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO: 20. In this embodiment of the invention it is particularly preferred that said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7-12 and 15-20.

[0155] In a further embodiment it is preferred that the said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 and 16.

[0156] In a further preferred variant of said embodiment said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 and 16 and at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:9-12 and 17-20.

[0157] Step e) of the method, namely the detection of the specific amplificates indicative of the methylation status of one or more CpG positions according to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 is carried out by means of real-time detection methods as described above.

[0158] Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, and complements thereof) without the need for pretreatment.

[0159] It is preferred that the methylation of one or more sequences selected from the group consisting SEQ ID NOS: 2, 3 and 4 are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the sequences of the group consisting SEQ ID NOS: 2, 3 and 4 are analysed.

[0160] In the first step of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, body fluids, stool or tissue embedded in paraffin. In the second step, the genomic DNA is extracted. This may be by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest

is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants e.g. by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA.

[0161] In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, in particular with methylation-sensitive restriction endonucleases.

[0162] In the third step, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

[0163] In the fourth step, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplificates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

[0164] In the fifth step the amplificates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridization analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

[0165] In the final step the presence or absence of colon cell proliferative disorder is deduced based upon the methylation state of at least one CpG dinucleotide sequence of SEQ ID NOS:1 to SEQ ID NO:4 & SEQ ID NOS:45 to SEQ ID NO:47, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NOS:1 to SEQ ID NO 4 and SEQ ID NOS:45 to SEQ ID NO:47.

[0166] In a further embodiment said deduction is based upon the methylation status of SEQ ID NO:2 only. In a further preferred embodiment said deduction is based upon the methylation status of SEQ ID 2 and one or more sequences chosen from, SEQ ID NOS:3 and 4.

Diagnostic Assays for Colon Cell Proliferative Disorders

[0167] The present invention enables diagnosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS: 45 to SEQ ID NO:47 may be used as markers. Said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for a diagnosis and/or prognosis of events which are disadvantageous to patients or individuals.

[0168] Specifically, the present invention provides for diagnostic cancer assays based on measurement of differential methylation of one or more CpG dinucleotide sequences of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, or of subregions thereof that comprise such a CpG dinucleotide sequence. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of one or more

CpG dinucleotide sequences of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO: 47 derived from the tissue sample, relative to a control sample, or a known standard and making a diagnosis or prognosis based thereon.

[0169] In particular preferred embodiments, inventive oligomers are used to assess the CpG dinucleotide methylation status, such as those based on SEQ ID NOS:1 to SEQ ID NOS:20 and 45 to 59, or arrays thereof, as well as in kits based thereon and useful for the diagnosis and/or prognosis of colon cell proliferative disorders.

Kits

[0170] Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent; a set of primer oligonucleotides containing at least two oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS:1 to SEQ ID NO:20 & SEQ ID NOS:45 to SEQ ID NO:59; oligonucleotides and/or PNA-oligomers; as well as instructions for carrying out and evaluating the described method.

[0171] More preferred is a kit comprising the oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS: 5-20.

[0172] Also preferred is a kit comprising the oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS: 7, 8, 15 & 16.

[0173] Further preferred is a kit comprising the oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS:7, 8, 15 and 16 and at least one of SEQ ID NOS:9-12 and 17-20.

[0174] In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE™, MSP, MethylLight™, HeavyMethyl™, COBRA™, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

[0175] Typical reagents (e.g., as might be found in a typical COBRA™-based kit) for COBRA™ analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit f or oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

[0176] Typical reagents (e.g., as might be found in a typical MethylLight®-based kit) for MethylLight® analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); Taq-Man® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

[0177] Typical reagents (e.g., as might be found in a typical Ms-SNuPE™-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for specific gene; reaction buffer (for the Ms-SNuPE™ reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

[0178] Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

[0179] While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following example serves only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

EXAMPLES

Material and Methods

Subjects for MethylLight Analysis

[0180] Colon tissues were obtained by surgical resection from 47 patients (29 male, 18 female) with colon cancer, with a median age of 66 years (range 31-93 years), from the tumor and a tumor-free location which was at least 2 cm distant from the tumor and which was confirmed to be without any tumor cell infiltration by histological assessment. In all 47 patients tissue samples from the colon cancer were obtained for molecular analysis, in 21 of these cases a matched non-cancer colon sample was also obtained for molecular analysis after tumor cell infiltration was ruled out by histological assessment. The metastatic lesions were obtained from 24 patients (13 male, 11 female, median age 64.5 yrs, range 41-79) with colorectal cancer that developed liver metastasis after prior successful colon cancer resection. In one case the primary colon cancer and a single liver metastasis were resected at the same time in a 74 year old female patient. Immediately after surgery, tissue samples were put in liquid nitrogen and stored at -80° C. until use. Formalin fixed tissues were processed as previously described and sections were stained with hematoxylin and eosin for histological evaluation. Tumor stages were assessed using the TNM-system.

DNA Extraction

[0181] Genomic DNA was extracted from the tissues using the proteinase K digestion method.

Genome-Wide Methylation Screening Assay

[0182] Differentially methylated genomic sequences were identified using Methylation-specific arbitrarily primed PCR analysis (MS AP-PCR) by comparison of different levels of disease to age-matched normal tissue for several different age groups (see for example Young J, Biden K G, Sinuns L A, Huggard P, Karamatic R, Eyre H J, Sutherland G R, Herath N, Barker M, Anderson G J, Fitzpatrick D R, Ramm G A, Jass J

R, Leggett B A. HPP1: a transmembrane protein-encoding gene commonly methylated in colorectal polyps and cancers. Proc Natl Acad Sci USA 2001; 98:265-270.) DNA from colon adenomas, CRC samples from patients with no lymph node involvement or metastasis and CRC samples from patient with lymph node involvement and/or metastasis was compared in MS AP-PCR experiments to age-matched normal colon tissue and to age-matched normal peripheral blood lymphocytes.

[0183] Five samples for each tissue type were collected for each of three age groups; over 65 years, 50 to 65 years and under 50 years. Genomic DNA was extracted from the tissue samples using Qiagen Genomic-Tip 500/G columns. The five DNA samples from each tissue type and each age group were pooled and experiments were performed as follows. DNA was digested with RsaI to generate smaller DNA fragments before digestion with MspI and HpaII, two restriction enzymes with different sensitivities to cytosine methylation. Then MS AP-PCR was performed as previously described by Liang et al. The digested DNA was amplified using these sets of primers: G1, 5'-GCGCCGACGT-3'; G5, 5'-TGC-GACGCC-3'; APBS5, 5'-CTCCCACGCG-3'. After amplification fragments were separated on polyacrylamide gels and those exhibiting a pattern of differential methylation were eluted from the gel, cloned into vectors and sequenced as outlined above. Identification of sequences was performed by BLAST searches in Genbank.

MethyLight™ Analysis

[0184] Genomic DNA was analyzed by the MethyLight technique after bisulfite conversion as previously reported by Eads et al. (Epigenetic patterns in the progression of esophageal adenocarcinoma. Cancer Res 2001; 61:3410-3418. and Fields of aberrant CpG island hypermethylation in Barrett's esophagus and associated adenocarcinoma. Cancer Res 2000; 60:5021-5026). In this analysis three oligos are used in every reaction. Two locus-specific PCR primers flank an oligonucleotide probe with a 5' fluorescent reporter dye (6FAM) and a 3' quencher dye (BHQ-1). For this analysis primers and probes are specifically designed to bind to bisulfite-converted DNA, which generally span 7 to 10 CpG dinucleotides. The gene of interest is then amplified and normalized to a reference set (β -actin (ACTB)) to normalize for input DNA. The specificity of the reactions for methylated DNA is confirmed using human sperm DNA (unmethylated) and CpGenome Universal Methylated DNA (Chemicon (subsidiary of Serologics) catalog #S7821) (methylated). For standardization the primers and the probe for analysis of the ACTB gene lack CpG dinucleotides so that amplification is possible regardless of methylation levels. TaqMan PCR reactions were performed in parallel with primers specific for the bisulfite-converted methylated sequence for a particular locus and with the ACTB reference primers. The ratio between the values was calculated in these two TaqMan analyses, using this approach the degree of methylation at that locus was determined. The extent of methylation at a specific locus was determined by the following formula:

$$\frac{[(\text{gene/actb})^{\text{sample}} : (\text{gene/actb})^{\text{SssI-treated genomic DNA}}] \times 100}{100}$$

A cut off value of 4% gave the best discrimination between normal and cancerous samples, as previously reported. Therefore, samples with $\geq 4\%$ fully methylated molecules were termed methylated, whereas samples with $< 4\%$ were

considered unmethylated. The primer and probe sequences are listed in Table 1 and were used as previously reported by Eads et al.

Bisulfite Sequencing

[0185] Bisulfite genomic sequencing was performed for the ALX4 gene in order to confirm the results obtained by MethyLight™ analysis. Briefly, bisulfite treated genomic DNA from 4 colon cancers and matched normal colon mucosa was amplified with primers specific for a fragment of the ALX4 gene containing 39 CpG sites and spanning the region that was analysed with the MethyLight™ assay: ALX4_bis1, 5'-TGAATAGGGTGATATTAGTTAGG-3'; ALX4_bis_2, 5'-ATAATCATCCCAAAACCTCTA-3' (SEQ ID NO: 60). PCR was carried out in a reaction mixture (25 μ l) containing 7 μ l of DNA, 0.2 mM dNTPs, 1 μ M primers, and 0.25 units of DyNAzyme EXT DNA Polymerase (Finnzymes). Amplification was performed using the following condition: 94° C. for 2 min, followed by 36 cycles (94° C. for 1 min, 45° C. for 1 min, 72° C. for 2 min) and then 72° C. for 10 minutes. PCR products were separated on 1% agarose gel, stained with ethidium bromide and visualized with an UV transilluminator. DNA fragments of interest were cloned into a plasmid vector with the TOPO TA cloning kit (Invitrogen, Carlsbad, Calif.) according to manufacturer's recommendations and sequence was confirmed by automated sequencing.

Statistical Analysis

[0186] The PMR values of the MethyLight assays were dichotomized for statistical purposes as previously reported by Eads et al. PMR values above 4% were considered as methylation positive and classified as '1', whereas PMR levels below 4% were classified as '0' (no methylation). This dichotomization should level off the quantitative impact of different levels of hypermethylation per gene, and allow the cross-gene comparison of methylation per gene in colon cancer and metastasis. The different clinicopathological features, such as location of primary tumor, grade of differentiation or stage of cancer were used as nominal variables in the Fisher's exact test or Chi square test. Otherwise Student's t-test was used to determine statistical difference. All tests were two-sided, and a p-value of < 0.05 was considered statistically significant.

Results

[0187] Differential methylation of a fragment of approximately 242 bp with genomic sequence matching a portion of the first intron of ALX4 was confirmed in seven different MS AP-PCR experiments comparing colon cancer and adenoma DNA to normal DNA obtained from the same individuals. In experiments performed on samples from patients over 65 years old, bands corresponding to a methylated fragment of ALX4 were found in DNA from a denomas, non-metastatic adenocarcinoma and metastatic adenocarcinoma samples when compared to normal age-matched colon tissue. The band was also identified in a mixture of pre-cancerous and adenocarcinoma samples compared to age-matched PBL DNA. An identical band was found in adenoma sample DNA from patients 50-65 years old when compared to normal age-matched colon tissue and a mixture of pre-cancerous and adenocarcinoma DNA from patients 50-65 years compared to age-matched PBL DNA. Lastly the fragment was found in a comparison of adenomas from patients less than 50 years compared to age-matched normal tissue (data not shown).

Confirmation of ALX4 Gene Methylation in Primary and Metastatic Colorectal Cancer

[0188] The methylation of the ALX4 gene was assessed in 47 colon cancers and compared to 21 cases of normal colon

mucosa which were obtained from a subset of these cancer patients. Using the Methylight assay a high degree of methylation was found in the cancerous colon as compared to the matched normal colon mucosa (FIG. 2A). Thirty cancers exhibited a PMR >4% (30/47) where as in none of the 21 normal colon samples ALX4 gene methylation was observed ($p<0.0001$). We then assessed the degree of ALX4 methylation in our series of metastatic samples and compared them to the primary colon cancers. In this series, apart from one patient, the tissues were, however, not matched and thus from different patients. Nonetheless, our analysis showed that a high degree of methylation can be found in both primary (30/47) and metastatic colon cancer (16/24), which did not differ by statistical analysis (FIG. 2B). Interestingly, primary colon cancer and metastatic tissues from the one female patient from which both tissues were obtained, exhibited a low degree of methylation in the primary cancer and a high degree of methylation in the metastatic lesion (FIG. 2B).

[0189] In order to confirm the results of ALX4 gene methylation obtained by Methylight assay we also performed bisulfite sequencing on 4 matched normal and colon cancer samples. These 4 patients were selected because there was a dramatic difference in the levels of methylation of the ALX4 gene in the cancer versus the matched normal colon sample (FIG. 3). The region of the ALX4 gene that was analysed by bisulfite sequencing spanned 39 CpG sites, including the 12 CpG sites that were analysed with the Methylight assay. The DNA fragments encoding ALX4 were amplified by PCR using bisulfite treated DNA and 3-11 clones per sample were sequenced. Similar to the results obtained with the Methylight assay we found that the majority of CpG sites were methylated in the cancer samples, whereas in the normal colon mucosa the CpG sites were widely unmethylated (FIG. 3). The varying degrees of methylation that were observed, however, in the cancer samples reflect the degree of contaminating non-malignant cells that were present in the DNA preparation from the cancer tissues used for methylation analysis. Overall, the results obtained by bisulfite sequencing confirmed the results of the PCR based assay, indicating that the Methylight assay correctly assesses the methylation of the ALX4 gene in colorectal cancer.

Analysis of Gene Methylation in Primary and Metastatic Colorectal Cancer

[0190] Next we analysed our tissues for gene methylation using a set of genes that had been previously reported by other groups to be associated with either colon cancer pathogenesis or the development of cancer metastasis: TPEF, p16/INK4A, APC, caveolin-2, DAPK, TIMP3. Using Methylight assays we assessed the methylation status of ALX4 and the other 6 genes in our series of primary and metastatic colon cancers. The cut-off of methylation was chosen to be a PMR (percentage of methylated reference) of >4% (as previously reported by Eads et al.) and all samples with a PMR >4% were classified as methylation positive ('1'), whereas samples with a PMR below 4% were considered methylation negative ('0'). The results of the methylation analysis of each gene in primary and colon cancer is given in table 3. In addition, the numbers were added giving the total numbers of methylated genes per sample. Using this approach we observed at least one methylated gene in 40 of 47 primary cancers, this would indicate a sensitivity of 85.1% for the detection of cancer (FIG. 4A). In contrast, only one of 21 normal colon mucosa samples exhibited TPEF gene methylation, indicating that the specificity of this marker set would be 95.2%. Thus, the analysis of only three genes, i.e. ALX4, TPEF and p16, allowed the identification of these 40 cancers,

whereas the other genes did not further contribute to the detection of colon cancer. In the metastatic lesions all 24 cases exhibited at least one of these methylated genes: ALX4, TPEF or APC, accounting for a detection rate of 100% (FIG. 4B).

ALX4 contributes to the identification of CIMP in Primary and Metastatic Colon Cancer

[0191] A condition termed CpG island methylator phenotype (CIMP) which is associated with microsatellite instability related to hMLH1 methylation, a proximal location in the colon and a family history of colon cancers was recently identified by Toyota et al. (CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999; 96:8681-8686). In our series CIMP+ cancers were observed in 20 of 47 cases (FIG. 4A). These cancers exhibited at least 3 methylated genes, which included ALX4, TPEF, APC, Caveolin-2, TIMP3 and p16. In the metastatic lesions we observed CIMP+ cancers in 9 of 24 cancers. The genes found to be methylated in CIMP+metastatic lesions included ALX4, TPEF, TIMP3 and APC (FIG. 4B). Colon cancers located in either the cecum, ascending colon, colon transversum or descending colon exhibited a CIMP+ status in 14 of 24 cases, whereas cancers of the sigma and/or rectum were CIMP+ in 6 of 22 cases, a difference that was also statistically significant ($p=0.026$). However, CTMP+ tumors were not associated with either early or advanced stages in our series, and showed no association with the degree of differentiation.

Association of Gene Methylation in Colon Cancer with Clinicopathological Features of Colon Cancer

[0192] In order to assess a potential association of the presence of methylation with the location of the primary tumor we classified our colorectal cancers into two groups: rectal cancers ($n=10$), and non-rectal cancers ($n=36$). Using Fisher's exact test, we found that there was a statistical significant difference in the presence of methylation with regard to the location of the primary tumor. While rectal cancers exhibited no methylation in 4 of the cases, whereas methylation of 1 to 7 genes was found in 6 cases, the vast majority of colon cancers (34/36) exhibited methylation in at least one gene ($p=0.014$). Thus, from this analysis we can assume that methylation is significantly more frequent in proximal, i.e. non-rectal cancers of the large intestine (FIG. 5A).

[0193] We analysed not only the association between the location of the primary and the overall presence of gene methylation per patient, but also analysed each single gene with regard to this association of location and gene methylation. However, apart from TPEF, none of the other genes—including ALX4—were linked to a certain location of the primary tumor, probably because of the low number of methylated genes observed in our study. However, TPEF was more frequently methylated in colon cancers (31/36) compared to rectal cancers (5/10) ($p=0.023$) (FIG. 5B). The total number of methylated genes in rectal cancers (19/70) was also compared to the total number of methylated genes in colon cancer (89/252), however, this difference did not reach statistical significance ($p=0.0513$).

[0194] While methylation is considered an early step in the process of colorectal cancer pathogenesis, we assessed the presence of methylation in the cancers with regard to the stage of the cancers. While none of our patients was in the UICC stage I, 20 patients presented with UICC stage II cancer, 13 with stage III and 13 with stage IV colorectal cancer. Interestingly, the number of patients without gene methylation increased with the UICC stage, in that patients with UICC stage II had at least 1 methylated gene compared to the more advanced stages, however this observation was not statistically significant. However, while we found no association

between the T stage and the presence of methylation, the presence of distant metastasis (M+ stage) was associated with significantly less methylated cancers compared to tumors without distant metastasis. Thus, while 3 of 6 cancers with distant metastasis did not show any gene methylation, cancers without distant metastasis (M0) presented with at least one methylated gene in 30 of 33 cases ($p=0.033$) (FIG. 5C). Again, the detailed analysis of every single gene with regard to the association with distant metastasis did not identify a single gene that was associated with the presence of distant metastasis, again maybe due to the limited number of cases analysed.

Discussion

[0195] ALX4 gene methylation was identified by MS AP-PCR in our study and we confirmed the presence of ALX4 methylation in a larger series of primary and metastatic colon cancers using the Methylight assay and bisulfite sequencing. ALX4 gene methylation was observed in 30 of 47 primary cancers and in none of the normal colon mucosa tissue samples. Furthermore, ALX4 was frequently methylated in the liver metastasis of a second set of patients with colorectal cancers. In the one female patient in which primary colon cancer and liver metastasis were resected at the same time we found a high degree of ALX4 gene methylation in the liver metastasis as opposed to the primary cancer, indicating that methylation of ALX4 may occur 'de novo' in the metastatic cancer cells even in the absence of methylation in the primary cancer cells.

[0196] In order to further assess the role of ALX4 gene methylation in primary and metastatic colon cancer, we analysed the presence of methylation of 6 other genes that have previously been linked either to colon cancer pathogenesis or metastatic development in our series of colon cancer and metastatic cancer tissues: TPEF/HPP1, p16/INK4A, APC, caveolin-2, DAPK and TIMP3. While TPEF and ALX4 exhibited a similar high frequency of methylation in primary colon cancer, p16 and APC were less frequently methylated. Of the 47 analysed primary colon cancers 40 exhibited at least one of the three methylated genes ALX4, TPEF or p16, indicating that a methylation based diagnostic test including these three genes may achieve a detection rate of 85.1%; since only one normal case exhibited a significant degree of methylation of the TPEF gene specificity would be 95.2%.

[0197] The cancers were then grouped according to the location of the primary cancer into rectal and non-rectal cancers. Overall gene methylation was more frequently present in non-rectal cancers compared to cancers of the rectum, an observation that has been reported by other groups as well. However, ALX4 gene methylation alone—in contrast to TPEF—was not associated with either location of the primary tumor, most probably due to the limited number of cases analysed in our study. Furthermore, since gene methylation has been considered to be an early event in the pathogenesis of this and other cancers, we assessed the frequency of methylation in each UICC stage and in cancers grouped according to the size of the primary tumor (T), presence of lymph nodes (N) or distant metastasis (M). However, colorectal cancers without distant metastasis frequently exhibited gene methylation, which contrasts with previous reports that gene hypermethylation may be associated with a poor prognosis in this and other cancers.

[0198] Recently Toyota et al. (CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999;

96:8681-8686.) reported the presence of multiple methylated genes in colon cancer, a condition they termed CpG island methylator phenotype (CIMP). In their analysis two types of methylation were identified: Type A methylation referred to the age-related methylation of CpG islands in tumors and the normal colon mucosa. In contrast, some genes exhibited methylation only in colon cancers which was classified as type C methylation. Interestingly, most CIMP+ cancers were associated with microsatellite instability related to hMLH1 methylation, a proximal location and a family history of colon cancers. In our series CIMP+ cancers were observed in 20 of 47 cases and were located primarily in the proximal colon cancer, an association which has already been reported by Toyota et al. and Rijnsoever et al. (CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999; 96:8681-8686. and Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. Gut 2002; 51:797-802). However, our CIMP+ tumors were not associated with either early or advanced stages in our series, which is similar to the findings reported by Toyota et al. In contrast to our study, Rijnsoever et al. reported that in their analysis CIMP+ colon cancers were poorly differentiated, a finding that was not observed in our and other studies and may be due to the larger patient group that was analysed in their study.

[0199] To our knowledge our study is the first to also address the role of gene methylation in metastatic colorectal cancers using a panel of 7 genes—including ALX4—that were analysed by the highly sensitive Methylight™ assay. Using this assay we found several genes to be methylated in both metastatic lesions and primary cancers, as well as genes that were neither methylated in metastasis nor in primary colorectal cancers. Apart from this observation, APC gene methylation increased in metastatic lesions compared to primary cancers. Based on these findings, we can classify the patterns of methylation in liver metastasis in three groups: class I genes: high degree of methylation in primary tumor and liver metastasis (ALX4, TPEF, p16), class II genes: higher degree of methylation in metastasis compared to primary tumor (APC) and class genes III: no methylation in either primary tumor or metastasis (caveolin-2, DAPK, TIMP3). Interestingly, all 24 metastases exhibited at least one methylated gene of the class I genes, indicating that this set of genes may be valuable for the methylation specific detection of liver metastasis in colon cancer. Interestingly, the two genes that exhibited a very high degree of methylation in the primary colon cancers, are also frequently methylated in the liver metastasis of colon cancer (ALX4, TPEF). From our analysis we assume that methylation of these genes occurs early in these cancers and that they remain methylated in the progression of the disease. However, the high frequency of methylation detected in the primary cancers and metastatic lesions makes them ideal candidates for a methylation-based diagnostic tool for localized and metastatic colorectal cancer.

[0200] In summary, using MS AP-PCR we identified the methylation of ALX4 in colorectal cancers and further analysis revealed that ALX4 gene methylation is a frequent event in colorectal cancer pathogenesis. Together with a further set of genes, ALX4 allows the identification of primary and metastatic colorectal cancers indicating that methylation based diagnostic tests may be helpful in the identification of this and other malignancies and, thus, may improve the detection and overall prognosis of patients with these cancers.

TABLE 1

List of primers and probes used for MethyLight analysis			
Gene	forward primer (5'-3')	reverse primer (5'-3')	probe sequence (5'-3')
ALX4	CGCGGTTTCGATTTAATGC	ACTCCGACTTAACCCGACGAT	6FAM-CGACGAAATTCTTAACGCAACCGCTTAA-BHQ1
Caveolin 2	TTTCGGATGGGAACCGGTGTA	CTCCCACCAGCGTTACC	6FAM-CCCGTCTAACCGTCCGCCCT-BHQ1
DAPK	TGCGTCGTTCGGTTAGTT	CCCTCCGAAACGCTATCGA	6FAM-CGACCATAAACGCCAACGCG-BHQ1
TPEF	TTTTTTTCGGACGTCGTTG	CCTCTACATACGCCGCGAAT	6FAM-AATTACCGAAAACATCGACCGA-BHQ1
p16/INK4A	TGGAATTTCGGTTGATTGGTT	AACAAACGTCGCACCTCCT	6FAM-ACCCGACCCGAACCGCG-BHQ1
APC	GAACCAAAACGCTCCCCAT	TTATATGTCGGTTACGTGCGTTATAT	6FAM-CCCGTCGAAAACCGCCGATTA-BHQ1
TIMP3	GCGTCGGAGGTTAAGGTTGTT	CTCTCCAAAATTACCGTACGCG	6FAM-AACTCGCTCGCCCGCAA-BHQ1
caveolin	TTTCGGATGGGAACCGGTGTA	CTCCCACCAGCGTTACC	6FAM-CCCGTCTAACCGTCCGCCCT-BHQ1

TABLE 2

Genes and sequences according to the invention.			
Gene name	Genomic SEQ ID NO	Methylated treated SEQ ID NOS:	Unmethylated treated SEQ ID NOS:
APC	1	5 & 6	13 & 14
ALX4	2	7 & 8	15 & 16
TPEF	3	9 & 10	17 & 18
p16	4	11 & 12	19 & 20
DAPK	45	48 & 49	54 & 55
TIMP3	46	50 & 51	56 & 57
Caveolin 2	47	52 & 53	58 & 59

TABLE 3

Summary of results from analysis of gene methylation in primary cancer and metastasis.				
Gene	Normal (n = 21)	Tumor (n = 47)	Metastasis (n = 24)	Class
ALX4	0/21	30/47*	16/24	I
TPEF	1/21	36/47*	19/24	I
p16	0/21	15/47#	6/24	I
APC	0/21	10/47#	10/24\$	II
TIMP3	1/21	11/47	2/24	III
DAPK	0/21	1/47	0/24	III
Caveolin 2	0/21	5/47	1/24	III

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 75

<210> SEQ ID NO 1
<211> LENGTH: 2470
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 1

```

aaagatgatt aaaagttaa ttgttcatct gaagagttga ttttttatt cctgtaataa      60
agggtacttt tagcagtctc tgctcatctt gcccacccgg ctcttttgc gggttgtaa     120
ggttataact tctgtgtctc agtaaacttg tgcatgccca ttttttctc tgttactacc    180
ttttctctta ttttgttta ttatggat gtaaaattac ctgttaattt tatttggaaat    240
gaaaaatttt aagggttcaca ttatccaat tctgtcaagat ccctacctct gtcatatgg    300
ttataatgtg ctgggtattt ttagacatgc ttataaaaa gatgtaaaac aaaataatga    360
tcactcctgt ggattttcc ttatggatgg agatgtctcc tttggctca ttacttctc    420
accccttgcc cattgatcag aggaggggtc ttaactatgg gtgaacccta tatcttactg   480
aagagggtt gttacatgta tatttcata atataactta catttacata gtacttttat    540

```

-continued

ttttagcata cctttttta ttaatccaa taatatcact gtaagttatg ttgaagcaga	600
ttgttaagtgt tcatttacaa atttgtaaat gaattaaaat gaaaggccaa agattaatc	660
atgaccaggc ctgaaattaa cacacaagac tcaattttt tcaaccaaag actttgttag	720
gtgatccctg cctgcaggac tcccccttct cctcagatgt cattggattt taccaggtt	780
actgttagatt ctagccgtt tagaactaac tagatctaag atgagtcccc tgatttcctt	840
tggttagagtc ttccaattgc tgaactccaa tattgtctgtg actagccagt gttacaacct	900
gtctgcctta ttttgtgtaa tggatttcat attacagagg catttttta atgtcaagat	960
gtttaagtat tgcttaagtg caaactactt aatactttt agctattaag taattaagat	1020
aggcaggatt ttatggttc caaaatgatt tgacctaaac taaaaagaga atgtggatct	1080
cctgaatctt acttgggttaa tcttaatata actcctagca ttctataatt cttccctaaag	1140
tcctcttacc tggctatctt ttgtatctt tttgtctctc ctctttttc ccagtcataa	1200
taactgcccag actctgttcc atttctctt gacagttct actcctaaagg tcatccattc	1260
tcttttaggtt tctttggcc tcagtttgag cacagcagat cccaaagacca catatgccat	1320
agcataggct attatagtc accttttga taaatgtat tgaactttat gtttagtaatt	1380
cttatttacc atcttcctat caaaaaggct taaagtcttc atttaatgct ctccctcatg	1440
tccatattgt taaaatgattt ccttttaatg acatcttaga acttcagaac tatttcacca	1500
tggaggatgt gtaagattag ccttttatca aataaaaagt gtgaaatgga atatgtatc	1560
tcattaatcc attctggctc taaaattctg tgactatcg ataaaattca gaaataaaat	1620
agtattacta atataaataa atttttatca taatttatatt tcctaagttt tgccgttaag	1680
aatgggtaaa atatctttaa aaccttgaag aaattttac ttgatagaaa gtttaatcca	1740
tctgtgagaa ggcaaatgtt ttcagacaca actaaagttc tctcttctat tttttttca	1800
tttatcttga actaagactc cactgtttca tcctctttaga tgctgtact tgaacaatat	1860
tgttttgaga caaaaacta gcatattaaac acaatttctc tttaaacgtct taagatttt	1920
gtttccctta cccctttttttaaaaaacaag cagccactaa attttttagt agtgaatttc	1980
aaaatccctt ttaacccattt aggtccaagg gtgccaagg atggctgcag cttcatatga	2040
tcaagttgtt aagcaagttt aggcactgaa gatggagaac tcaaatcttc gacaagagct	2100
agaagataat tccaaatcata ttacaaaact ggaaactgag gcatctata tgaaggtatc	2160
aagactgtga cttttatattt tagtttatcc atttttatttcc agtattccct cttgtaaact	2220
tgaggtttaga cactttactt aaaagtgtat tttaaattaa gcaataatata ttaaaacttt	2280
tcttgcaaaa gtttagcattt atatttttaa ataagatata ttgaattcat tcaagtgtatc	2340
atataaagaa aataagtgtt aaaaactccat ggctgttag ttcttagttc ttttttaagat	2400
taaagagaag agaccaaata tagcatcact gtactgaggc aaggtttct gtgttagttca	2460
tagaaaaacttag	2470

<210> SEQ ID NO 2
<211> LENGTH: 2229
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE : 2

tcttcctcg gcgctggctg gtgcgggttg gggtcaggtg gagaagccgc tctttgttaa 60

-continued

ggtgacagaa cgtgctgggg gtggggccg gggccaggc cggtgcaact agggggccgc	120
tgcctttcc tggacacagt ggaagcttct tcgcacatcac caaattttg tcataccttc	180
tgagggacct gcttcaggc agcacgcaag ttgttgtccc gggtttactc cgccccctc	240
tactgggtga ggaaggagca tcttgaatgg agatgggggt gtcccccgtt tatacatctg	300
cagagaagag gtgtgccggg ctgcacactc ggaggccgcg gtaactgata ttagagaaga	360
ccccgggtgc agctggaaag gctcaactggc tggaaagagg tgccctctcc ttccagcaaa	420
gggcctgtt tggaaaggct gcttctcacc tgtcttagtgg caccacagga cggtcggcct	480
ccactcgatat tccccggac ggttatcatca catagccggg tcctcgaagt gttggttcc	540
caatccgatg actgtcacct cggtgaggac ctgtgtgtat ggccggagaa ccctgcgctg	600
cgggcgcaca tggccaggtg ggcctggca ggacgtcc gggtgcagga cggcgcctt	660
accggccccac cccaaaccgt tgcctggcc taggtccttc ggcttcgttga acaggggttt	720
ggggggctaa ggacgcttag gtcgggggg caggaagtgc tctctggta agcgttctc	780
cttctctccg gcatacactc ccctaccac ccacctcgcc taccctcggtt gcgagaggct	840
ccaacaaggca gggcgccccc ccccatgaa tcatcccaag gcctctgagc cgcggggct	900
cggggcaact atccccctcc tctctggcc tcaggcaccc cagtcacggg gtctgcagag	960
aagccccgaag cccggacaaa cgccggcggac gtcaacaacc tctcatccct ggcagoagca	1020
aaggccaata tatttccatt tcttatttca gtttgcacc accaaaaaagc tgcgcgcggc	1080
tgagggcagg aaggcgttaga gaccgagaag aagggacgtc cggagaaag tgcgcgcagg	1140
tgatcttaga aaccagagtc ctccggact tgcggagat tttctgttagg gcttttaat	1200
ctgtttctct actgcgtgcc gggtcgcaag cgcgtcgccgc tcagggtttt gtgactccgg	1260
cttagccccgg cggtcgcggc gaggttcctg ggcggccgc ttggaaacttc gcattagaat	1320
cgggaccgcg caaatgcctt gggtgaatgt tcaccctatt caagaaacac tgctgtcagg	1380
aacaaaatgg ggtccccggt gctccgaatgt atcttctgaa attttcttaa aacaacttac	1440
aaaaaatgtt tttgttttaa cgtttacaa cgtttaagga aacatgtaaa tggctgtttt	1500
ctttatcgat atggcgtcc taactaacag tgtacacata cataacaatt cttccaactt	1560
tcctcctcag agctaaggcac ttcaactatata gtaaaattata ataaagaaaa gattgtgcaa	1620
gatcatgcaa gtgcattgac taaaatatt gagtttaat ccagggccctc tgttttctta	1680
ttaacaact tttgtgtttt gaccagactg gtgaagcagg ctatggaaat taacaaagta	1740
aaaaattaaa agcatctcc ttgcacatcc ctccctccaa aattaaacaa cagtcgcccc	1800
ttcctgagca ggttcgttc ccaggctcgat gtttctgc gatcacccca cagtcacccca	1860
cagcagctgt tgcgttttgc gtcgggtttt cgtttctgat ttctttgggt cgtctttgt	1920
atacaaaaca caccccaatgt ctctaaactaa attcaaatac gacccggca gaatttacac	1980
atttcgttgtt gcatggattt tgtcgggtca gggaaataa ataccctgt gtatataacc	2040
actgagtcata attcgaaaaa tggggactgg gcccctaggc ggcacccca gggctccaac	2100
ctggccccggc cctccccaga ctttggcgct gagagcgctg cttttgggg tgggtggacg	2160
gagaggtaac aatctgcttt caacaaaaac ctgtcgccac cgaatcgaaa gcgaaaggga	2220
agggagaag	2229

-continued

<210> SEQ ID NO 3
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 3

gtctttggtg agatatgtgt tttacaagtt ttaatggaga aaaatgtaa	60
cctgaaactt ggctatattga gtaatgagaa aatagtca	120
tcccccagga cagtggttct	
caatcatggc tatgtgtttc tccaggaaaa cttaaaaat atataatatac caatgtttct	180
gtgtcacttc tagggattcc aagtcttga atacgaactc tgcatcagta ttcttaatt	240
atccaggtga ttgtgatgtg aaatcatgac tgagccccac tgctctaaga tgaaataaac	300
ttcctcagc actgaaatca caaacttaaa ctaccaaaat taattaaggg catggaaatc	360
aataaggcat agggaaagtt ttacattata aaatttatttc tttaaatcac agctcattgt	420
ttatatgtta ttgccattt tagaaaaggg tgaaaaataa gcaaatttaa ttactctcag	480
tttgaaaaat tatccagaaa tgaagatgac gactctgaaa cattgtcaat atcatttgac	540
ctataaataa tggtaataa catttactac acactgata	600
atcttatataa aactaaggca ataatgcatt tagaacatc tatctatatac tatgtatctt	660
aaggtaggcta gaaattaaga tatgagttat taagtatgag atgttaaggt gtggggtag	720
aaatttataact gtacttcatt atcaataatc aacatataact tcaatatcac atacatttaa	780
ctttaatttg tacatcttta actatttta attatgtgtta taaatataag tacacacatc	840
tttatgtatt tatttattca tacctccatt cacttattta tataggggat cccccc当地	900
ccactaccat taaaccatac atttttattt taatctttag aacaagccca ggaggoaggt	960
attgttattt ctcacatccc acaaatttgc gactcataa gttactgtgt	1020
cagacatatt agaagcttaa tacatatttgc gtgaacatatac gcataaaaaac agagagacag	1080
acatgtacaa cagctcatct ttacactgat taaaagctt taacctgtct cagaaacctc	1140
tctgtgaaaa ctgagcaaaa atcgaggat ctttcattt gtcataatagg tataagggt	1200
accttacttc tccaacaagg atgaatattt aatgtggat cccaggcccc aactccagat	1260
tttctgaatc cctgatagtg ggacttggaa tttgtctatt gttcaatgt ttctcaagga	1320
attcatatga tcaaccaggc tcagaaatca ctggatctt ttgccgaatg ttgagaatta	1380
aagtttgggc cttactgegg ctccacagaa agggcaatg aagtatcatg gacagaactg	1440
atacgttccc agttatgttc ccctctcaga agctaaccagg cagcaatatac gcagaaat	1500
gtgactttagt tcttgtgtc tgaagtgcagg cagaatttca cagactccca gcagtgtcac	1560
tgacgagatt tttttcttgg ggcaagttgc ctgtatgttt caaagccata ttccctttat	1620
ataaaaatgag ataatatttct ttgtctcata ggggtgtttt aaagattaaa taaaataac	1680
atgttctatc ctacatggca caatgcgtca caccttgc gcaaaggata catcttac	1740
tatttgaagc aatcagaaatca tgaaggat aagagttctg attggcagtg	1800
tatcttattt tccagggttc atttatttctt cttaaactat tcttggatggaa atataactcc	1860
caagccccctt acttaagctg tgagtaatctt cacactttat aatgtatgttc ttccatgag	1920
aaaaaaaaat gttcttaagt tttctggaga aaatataatctt gcactatttc tactgaaaaaa	1980
tctaaacaact ggactctgtc cctctgcatac aattcttagag tggatgtgcc acaaataaag	2040
tgttcttagt caagaagattt gaaagttaaat atggtatagt atttttttttt aagaattttt	2100

-continued

caaatacatg	gtatgattgt	gtcatattac	tagcaatcat	atgatacgca	atgcaaagta	2160
cagttcatag	acttaaattt	aattctaata	agtaaactga	ttttgccttg	ctggggaaaa	2220
gttaaagcac	taatccaatt	gctaattgcag	tcttgtctac	ttctttggta	cctagtgaca	2280
agtctaaata	atgtatataat	ttttatttac	atattcagta	atacaattct	ctgctcaatg	2340
agtgtatgttc	ttctgccact	tggtgggtgc	tgccagttc	agaatttgc	tcttggtggc	2400
actataaacac	taagtacaga	gtaagtgc当地	caaaatttgc当地	gcattcccat	tgaaaaggct	2460
ttgcttcaaa	ctgtttaata	atttaagga	cctctgtgga	agcaaccgca	tttgtttaacc	2520
agttacaacc	agtaattaac	tcctttggag	ttttactta	cttttggc当地	aacgtcttag	2580
gaagagcata	tattattaga	aagtatgc当地	aaaatttact	tagcagaaaa	ttcaaaaaca	2640
gtttccctct	gctaagaggt	tctctaaat	tctacttaca	tagccaaact	ctgaaatcct	2700
agcagggtct	gtttcattat	cataattact	gcataaaacac	ttttaaggac	tttgccttta	2760
gtttcaagca	tgacttattt	tcataagcct	gattagttac	cacaccagcc	ttgctatgga	2820
aatgacatg	tttcatttct	ctgctgtaga	gttgttaat	cttgc当地tat	atttatgttgc	2880
ccttcctgc	tgaaagcctg	tagc当地aaaga	aatttctaaat	tccttgc当地ttt	gcaatattag	2940
ttggcagctc	tatctaatgg	gtattctgtt	tccttaaaga	atttagctgc	tctgtctaga	3000
agccgatttt	ctgatgc当地tctc	caacgtctgg	tctaaattgat	ctgttttaat	ggagtttgc当地	3060
tcggtgagga	gcgagatgc当地	accgactaga	atgctggat	ctgctgctta	attgccc当地gga	3120
gtggagagaca	ctgagattca	gaaatctttg	gaggtggag	gggagagggg	cagtctcgga	3180
cgaggcgga	gatgttaagat	aaagggatgg	atttcacaca	ggaaaaaaaa	aaagatttgc当地	3240
ttgaggcact	gaggtgctgc当地	acgatcaca	ctctcaaagg	agaagttaaa	aaagcaaggaa	3300
gtgggaggag	gttggagggt	aaagtactta	aaaggattac	tcgggtacaa	tttgc当地ttt	3360
tgctggtgctc	tgcaaaggat	agatagtccc	gttttcaaag	tatatgaatg	cctcttttaa	3420
gtgattggga	atggacacta	attgc当地tgtt	aaatgttatac	aatgtctc当地	ctaaattcag	3480
gggacacaga	aagaggggca	caaaaggaga	attnaaatag	aaaaaggag	gatccggagg	3540
ctttgaaag	cggggggaga	agaaggagga	gggataacag	agaggaatag	agaaggagag	3600
cggagagaag	ataaacaaaa	acaaaaacac	gaatcactga	ataatcacac	accaaaaaa	3660
aagetcttcc	ctatgggca	tccaaaacac	tgagactgca	atagtgc当地cc	cggtcatgga	3720
agaaagatgt	tcctctccac	ccttgc当地cc	gaaagctt	ggtcccgta	ctggcgacta	3780
aaattccatt	aggctaaaga	gtgtgtctaa	ctgc当地ctgaag	aatgc当地cg	acggaaggcg	3840
ggtcccgctc	tgccgcttgc当地	ccttcccgct	ggagagaatg	aaagaaaacgc	gcagagccag	3900
agactcctgc当地	cgagtttagac	cttctctcg	cgccccaggt	caccggccat	ccggcaaaga	3960
cccgagtaag	gaacgc当地ggg	tcactgc当地tgc当地	ggccaaacaaa	tggagccgc当地	tctccctc当地	4020
ccggacgccc	ctgccc当地ggcc	gatgc当地cc	gcaacccacc	cgccggegtat	gcagaggagc	4080
ctttctcttt	ctctcagacc	acttgc当地cc	accaatctga	ccttccaaac	acatctgacc	4140
gcacctccca	ggtggacaca	ctaataaggct	acgggctgga	gaggagccgg	tgtatgaggag	4200
agggattcaa	acctgc当地gaac	gcttggctg	ggtc当地ggagct	gcggggggcc	tgggaggaga	4260
gagggaggaa	gagagaagga	aggagagcgc当地	ctgccc当地ggat	ggctgagctg	cctcggc当地gag	4320
cagccttggg	gttgc当地cgct	cttgc当地ggag	atgctgctgt	tgcttccagg	tcggcaagag	4380

-continued

cggttctaaccatcgcccttcacccctttccctgtaaa	tccctagaga	aacgtccctg	4440
gcctctccgcgcgacattccagcctgca	tccccctaca	gcctaggcgg	4500
cacgctggagcgccggtcgc	cagcaggac	ccctctcccg	4560
ccctgctgtctgtgtctc	tgacaccc	gcccccacca	4620
cacccagccgcggccgcac	tcggggcccc	gggtcacgcg	4680
ggaccccgccatcggtaggg	ggcgtccctgg	agcggagagt	4740
gtgcgggtagcccaccctga	agcccgagct	tctcatttga	4800
actcggggcca	gcgcctggcg	actgtggcttccgc	4860
cattttgcacccactcg	gaccctccc	tcccgggacc	4920
agagcccgac	tgcgttcc	cacgtgaccc	4980
acggccagagg	gatgcggcg	gcagagctcg	5040
gactctccctccaccctggcc	ccactcgct	gccccctggac	5100
tcttgttctccggttccca	gagctccctc	tttatggcag	5160
gcagcttc	agcggacac	cctctcgctc	5220
gaaactctcg	agatcatcg	cgggtttggc	5280
cggccgcgc	tctgtgtcg	ccgtccgggg	5340
cgatccctgttccctggaa	gcgcgttgc	gctgcagagt	5400
ctgtgggagtccccggcga	gtcagcgc	tggacactt	5460
ctgtgtgtccgtcatgt	actcatcgta	gcccccccg	5520
accccttaatgtactgcca	aacgcccacc	ggctggaaatt	5580
cccgcccccgacccttaac	tcggcagaag	aacacgctgt	5640
cctagcgcgc	cctctcgac	cctcaccc	5700
tctatccaga	gaaacaagg	gaactgttgc	5760
ggcagaaag	tgcagggtta	gcaggaaacc	5820
gaggatttt	aacggatgg	tttctgggt	5880
caatcctagg	ggaaacact	aaacagaagc	5940
aggatgagga	agaaataact	ttatgagaaa	6000
tgactgcagg	ggaatcgctg	attcctggca	6060
ttgaagacca	ggtcacacag	attctagagg	6120
tctctctctc	tctctctctc	tctctctatc	6180
tcattccctt	cttccttagg	cggcaaaaga	6240
tctttgttc	cctaagcttc	aaggtagtac	6300
gtctcccaagg	ccggagtg	catgactgag	6360
ggcggAACG	acttcagtgg	gcgcgtgact	6420
ggtcaggcac	tcttcttgct	cgctcggtt	6480
ggcttcgc	tgtcgaagct	tcctaattaa	6540
cctagctggc	atagcattcg	gttgacagc	6600
cttctgtgaa	tataatttt	ctgtagttaa	6660
		atctggctct	
		gaataaagt	
		tctttcaaag	

-continued

atgtatataa gctgaagtgt atgtaacttt agagaggagg gaatgaccaa ctgtaactca	6720
gggtgaaagc ctgtatagtt cctagttatt actgatgtaa atgc当地aaag gaaaattatt	6780
atgc当地tattc ctaatttatc ctttacaag acaagtttag atatgcaacc ctattagatt	6840
tgggtcaata gattgttctc tttttggca gtttctaaat ttggcathtt aataaaactc	6900
aacatgttc tataacttct tgattcatgc gtacatgtgt gttgttttg aaagaataag	6960
tttcactttg ctatgccta atcactttt agatgcttta ttatgtaat aattatgagc	7020
ctgcaaaaac aatttttggaa aatgttgatg gctttgtagt ccaacacaga ctggttgct	7080
tcattcctag cccttgcatt gttttaggaa ataactaact taaatgtgaa gttgacattt	7140
gcaatcaaga aattacatata ttaccagata ttttaaaggg gactgcataa actaaagaga	7200
ataaaactggt tttcagata ggttgtcaag aacttggcac cccgcttcca cccctgttaa	7260
ctttagaggta atcaatcttc atttgagcca aacagaccat cacagaaaaac actgtgcctg	7320
tttatcttta ttatgaggc tttgttctt ctttgtctgg atacattca aataagggtt	7380
tgtttcagtc gttgaagcaa aagaacaatt aaagatgggg aaatggtaaa agggattca	7440
gagatcatca ctatgtctt tccaaaatgt ggagtttgc ggtcataat attgtccacc	7500
taatgagcaa aaaataaaaaa taaaaaaaaa acaggaagca aatgttaagc tttcattcac	7560
cactgtcagt attaacgcaa gctttaaaaa atagcactat cagaaaaagg tactaaagga	7620
gaattgacta gaaaagaatt gtggaaaatg gaaacgaata ttgatcactt aactagattt	7680
tgaggttatac agtagacagt gaccttgcag tacagctata gttgttgat ttttttttta	7740
ggacaagttt tttaaagttt caaaatgttg ctttttttt tttttttttt gtaagatgtt	7800
ttaatgactg gagtgttctc tttgaatttg agg	7833

<210> SEQ ID NO 4
<211> LENGTH: 5666
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 4

aaaatttagaa cttttaccc cttgcgtttt ttataactttt tagtgetgtt taacttttct	60
ttgttaagtga ggggtgggaa ggggtggccat aatctttca gggagtaagt tcttcttgg	120
ctttctttctt ttctttctt ctttttctt tgagaccaag ttgcgtttt gtctccagg	180
ctggagtgcata atggcgccat ctggcgctac tgcaacccccc gccttcttgc ggggttcaagc	240
gattctccata catcagccctc cgagtagctg ggattacagg catgcgcac caagccccgc	300
taatttgtat ttttttagta gagacagggtt ttcgcattgt tggtcagggt tgcgttgcac	360
tcctggccctc aggtgateccg cttgtctcgg cttcccaagaa tgctggattt atagacgtga	420
gccaccgcatttcc cttttatgtat atagtataaa ttctatccaa agcattttt	480
tttttttttgc agtcggagtc tcattctgtc acccaggctg gagggtgggt ggcgcgttcc	540
ggcttactgc aacctctgc tcccggttcc aagcgattt cctgcctcgc cttctgtgt	600
agctgaaattt acacacgtgc gccaccatgg ccagcttattt tttgtatattt tagtagagac	660
gggggtgtcac cattttggcc aagctggccat cgaactccctg acctcagggtt atctgcccgc	720
ctcggcttcc ctttttttttgc aagctgtcg ggattacagg tgcgttgcac cgcgttgc tccaaagcat	780
tttctttcttgc tgcctcaaaa caagattgca agccagtcct ctttttttttgc aattcaagag	840

-continued

ctaacaggta ttagctttagg atgtgtggca ctgttcttaa ggcttatatg tattaataca	900
tcatttaaac tcacaacaac ccctataaag cagggggcac tcatattccc ttccccctt	960
ataattacga aaaatgcaag gtattttcag taggaaagag aaatgtgaga agtgtgaagg	1020
agacaggaca gtatgttgcag ctggcttttgc gatcactgtg caactctgtc tctagaacac	1080
ttagacttt ttctggctta ggaattatga ctttgagaat ggagtccgtc cttccaatga	1140
ctccctcccc atttccttat ctgcctacag gcagaattct ccccccgtccg tattaataaa	1200
acctcatctt ttcagagtct gctcttatac caggcaatgt acacgtctga gaaacccttg	1260
ccccagacag ccgttttaca cgccaggagg gaaggggggg ggaaggagag agcagtccga	1320
ctctccaaaa ggaatccttt gaacttagggt ttctgactta gtgaaccccg cgctccgtaa	1380
aatcaagggt tgagggggta gggggacact ttcttagtgc acagggtgatt tcgatttcg	1440
gtggggctct cacaactagg aaagaatagt ttgcctttt cttatgatta aaagaagaag	1500
ccatacttcc cctatgacac caaacacccc gattcaattt ggcagttagg aaggttgtat	1560
cgcggaggaa ggaaacgggg cggggggcgga tttctttta acagagtcaa cgcactcaaa	1620
cacgcctttt ctggcaggcg ggggagcgcg gctggggagca gggaggccgg aggggggtgt	1680
ggggggcagg tggggaggag cccagtcctc cttccttgcc aacgctggct ctggcgagg	1740
ctgttccgg ctggtgcggc cggggggagac ccaacctggg ggcacttcag ggggtgcaca	1800
ttcgctaagt gctcggagtt aatagcacct cctccgagca ctcgctcacg ggcgtccctt	1860
gcctggaaag ataccgcggt ccotccagag gatttgggg acagggtcgg agggggctct	1920
tccggcagca cccggaggaaag aaagaggagg ggctggctgg tcaccagagg gtggggcgga	1980
ccgcgtgcgc tcggcggctg cggagaggggg gagagcaggc agcggggccgc gggagcagc	2040
atggagccgg cggggggggag cagcatggag cttcggctg actggctggc cacggccgc	2100
ggccggggtc gggtagagga ggtgcggggc ctgctggagg cggggggcgct gcccacgc	2160
ccgaatagtt acggtcggag cccgateccag gtgggttagag ggtctgcagc gggagcagg	2220
gatggccggc gactctggag gacgaagttt gcagggaaat tggaaatcagg tagcgttgc	2280
attctccgga aaaagggggag gtttcctggg gagtttcag aagggttttgc taatcacaga	2340
cctcctctgc gcgacgcctt gggggcttgg gaagccaaagg aagaggaatg aggagccacg	2400
cgcgtacaga tctctcgaaat gctgagaaga tctgaagggg ggaacatatt tgtattagat	2460
ggaagtatgc tctttatcg atacaaaatt tacgaacgtt tggataaaa agggagtctt	2520
aaagaaatgt aagatgtgtc gggactactt agcctccaaat tcacagatac ctggatggag	2580
cttatcttcc ttacttaggag ggattatcg tggaaatctg tgggtgtatgt tggaaataat	2640
atcgaatata aattttgc gaaattatttc agaagcggcc gggcgegggtg cctcacgc	2700
tgtatccct tcactttggg agatcaaggc gggggaaatc acctgaggtc gggagttcg	2760
gaccagcctg gccaacaggt gaaacctcgct ctctactaaa aataaaaaa gtagccgggg	2820
gtggtggcag ggcgcgttaa tcccagctac tcggggagttt gaggcaggag aatcggttgc	2880
acccggggagg ctgaggttgt agtgaacagc gagatggagc cacttcactc cagcctgggt	2940
gacagagtga gactttgtcg aaagaaagaa agagagaaag agagagagaa aaattattca	3000
gaagcaacta catattgtgt ttatTTTaa ctgagtaggg caaataaata tatgtttgt	3060
gttaggaactt aggaataat gagccacatt catgtgatca ttccagaggt aatatgtgt	3120

-continued

taccatttt ggaatatctg ctaacatttt tgctctttta ctatcttag cttacttgat	3180
atagtttatt tgtgataaga gtttcaatt ctcattttt gaacagaggt gtttccctc	3240
tccctactcc tgtttgtga gggagttagg ggaggattt aaagtaatta atacatgggt	3300
aacttagcat ctctaaaatt ttgccaaacag cttgaaccgg ggagttggc tttgttagtcc	3360
tacaatatct tagaagagac cttatgtt taaaaacaaa aaggaaaaag aaaagtggat	3420
agtttgcaca atttttaatg gagaaggagg aagaacatgt agaaaagggg aaatgtatgtt	3480
ggcttagaat cctaactaca ttggtgttta atataggaac atttatttat ataacatttt	3540
aaagtactaa attcatatta gtatattatc aaatggatattt attatcaat gggtttaagc	3600
atcctacaca tttaattca attgattcat ttcttttg ctggattt ctatcatgat	3660
ttaaatattt acatatgggt tacttttag atttttcata ctatgaaata taagaaaaac	3720
ctttaaggct agtttatga ccaagacgaa ggacttcattt gaatacacaa aacaataat	3780
atactgcac acattttgtt tctttttgtt gctgcaattt gggttgcata tactttctct	3840
ttgtctcttt gaaaactgag tcagttcac tttctcagga caggattta taaccataat	3900
ataattttagt ataattccctt gatttaggca aattatgcaat tttgtgttta gtatgaaatg	3960
tacctaaaaa taagtaactc ctcttaaca ccaccatcct caaactaata taacaataaa	4020
cagttatcctt aaaataaattt gtctacttcc accatgcagc actcaaattt taaggttgct	4080
atgactgcag acagtattttt aaaattccctc tctggaaatg gctttgttcc caagatgatt	4140
taggaaccaa agaggtgacc atcttctgtt taatgaaactc tcaaatacata aacctggaa	4200
gtgttttagt ttctactgc tgctgttaca aattatcaca aatgtgttag ctAAAacaaa	4260
cacaaaatta ttatTTACA gttctagaga tcagaagtca aaaatgggtt cacaagggtt	4320
cattttttt ggaaactcta aggggcaatc tgtttcttg tttttccag ttcttagtga	4380
ccatcaaattt cctggctca tggctctgtt attttcttg tggcctgtgc ttccattttt	4440
gtatcttc tctgactgtg accctctaattt aaaaacactt ggggttatgt tgggeccacc	4500
ctgaaaattt tggataatctt ccctcaagac catatattaa atcacatctg caaagctct	4560
tttgcacat aagttaatgtt attaaaagtt tttgaggattt aggacataga cattgggggt	4620
gggggggcat tattcagctt accacaggaa ggaatttttag ggttaattaa actagcttc	4680
ttatTTATA cttgaagaaa ttgaagttt ggaattggag agcattatgc taaatgaaat	4740
aagccaaaca cagaaagaca aatatcacat gttctactt atctgtgaaa tataaaacaa	4800
ttacattctt agcagtaaag agtagaatgg tggtaacttag agctgggggg tgggaggaat	4860
ggggagatgg taatcaagat ataaagecctc agttaagatg ggaggataaa gtttgattgt	4920
tttttttagt atgtgtttca tagcatgatg aatatacgta aatagtaat cccaaatgt	4980
ctcatttgac aaaaatgtca aatatttgatg atgatggata ggttaacttag ctgtacttac	5040
taattccccca ttgtgttcaa agatcataac ttcatattgtt accacataaa tatatacaac	5100
tgtactatcc caatataaa tttaaaaactt aatataatga aaaagaaaattt gaagttcaac	5160
atccccagaa gctaagtgtt aactttttttt tttgtgagaa tttgtttttaa caaacaaaaca	5220
agttttctct tttaacaat taccacattc tgcgttgaa tatacagcag tgaacaaaaaa	5280
aaaaaaaaaaa aaaaaaaaaatc tccaggccta acataatttc aggaagaaat ttcagtagtt	5340
gtatctcagg ggaataacag gaagttagcc tggagtaaaa gtcagttgtt ccctgcctt	5400

-continued

ttgtctatcc gccccgtcgtt cacagtgttc tctgcgttgtt acggacagtc cggcagaaggtt	5460
cggaggatataatggaaattc atttgtgtactt gaagaatggat tagagaactca aagaaggaaa	5520
ttggaaaatcg gaagcaaattt taggggttaat tagacacccgtt gggcttgggtgtt gggggcttc	5580
ttggcggtgatgggggctcta cacaaggcttc ctggcggtca tgccggcccc accctggct	5640
ctggaccatttttttggcagg	5666

<210> SEQ ID NO 5
<211> LENGTH: 2470
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 5

aaagatgatt aaaagttaa ttgttattt gaagagttga ttttttatt tttgtataaa 60
agggtatTTT tagtagttt tgTTTattt gTTattcgg tttttttgt ggTTgtAA 120
ggTTataatt ttTgtgtTT agTaaatttT tGTatgtta ttttttttt tgTTatttt 180
tttttttttA ttttgTTtttA ttatTTgt ataaaatttA ttgttaattt tatttgaat 240
gaaaaatttt aaggTTtata ttatTTaaAT ttTgttagat ttttattttt gttatATggT 300
ttataatgtg ttgggtattt tttagatttG ttatTTAAA gatgtAAAat aaataatgaa 360
ttatTTTTgt ggatTTTTT ttatTTTTg agatTTTTT ttgggtgtA ttatTTTTT 420
atTTTTgtt tattgattAG aggaggggTT ttaattatgg gtGAattttA tatttattg 480
aaagaggTTt atttatgtA tattttata atataatttA tattttataA gtatTTTT 540
tttttagtata ttTTTTtA ttaatttAA taatatttA gtaaggTTA ttgaagttagA 600
ttgtaaagtgt ttatTTataA attgtgAAat gaattttAAat gaaagggtAA agattAAatt 660
atgatttaggt ttgaaattAA tataAAagat ttaatttttA ttaatttAAag atTTTTgtA 720
gtgatTTTTg ttTgttaggt ttTTTTTTt ttttagatgt tattggattg tatttaggtt 780
attgttagatt tttagtcgtg tagAAatttA tagatttAA agatTTTTT tgatTTTTT 840
tggtagagt tttaatttG tgaattttA tattgtcgTt agatTTtagt gttataattt 900
gtttgttttA ttttgtgtAA tggattttt attatAGgg ttttttttA atgttaagat 960
gtttaagtat tggTTaagtG taaatttAAt aatTTTTT agtttattAAG taatttAAG 1020
aggtaggatt ttatTTgtt taaatgatt tgatTTAAat taaaAGAGA atgtggattt 1080
tttgaatttA atttgggtAA tttaatata atttttagtA ttttataatt ttttttAAAG 1140
ttttttattt tggTTatttA ttgtattttA ttggTTTTT tttttttttt ttagttataA 1200
taattgttag atTTTTgtt attttttttt gatagTTTTt atttttaagg ttatttattt 1260
tttttaggtA ttTTTTGGTT tttagTTGAG tatagttagat tttaagattA tatagtttat 1320
agtatAGgtt attatAGttA atttttGAA taaatgtgtA tgaattttt gtttagtAatt 1380
tttatttattt atttttttat taaaAGGtt taaAGTTTTt atttaatgtt tttttttatG 1440
tttattttgt taaatgattG ttTTTAatG atatTTtagA atttttagAat tattttattA 1500
tggaggatgt gtaaggattA ttTTTTtAAt aaaaaAGGt gtaAAatGGA atatgtAatt 1560
ttattaattt attttggTTt taaaattttG tgattattAG ataaaatttA gaaataaaaat 1620
agtatttattA atataAAtAA atttttattA taatttattA ttttaagttt gttttgtAAG 1680

-continued

aatgggttaaa atattttaa aatttgaaag aaatttattat ttgatagaaa gtttaattta	1740
tttggatgaa ggttaatgtt ttttagatata attaaagttt ttttttttat tttatattta	1800
tttattttga attaagatttatttttta tttttttaga tggtttattt tgaataatata	1860
tgttttgaga ttaaaaatttta gtatattaat ataattttttt ttaaacgttt taagatttt	1920
gtttttttta tttttttttt taaaataag tagttattaa atttttttagt agtgaatttt	1980
aaaattttttt ttaatttttat agggttaagg gtagttaagg atgggtttag ttttatatga	2040
tttagttgtta aagtaagtgtt aggttattgaa gatggagaat ttaaattttc gataagagtt	2100
agaagataat tttaatttttatt ttataaaattt gggaaattttag gtatattaata tgaaggttatt	2160
aagatttgta ttttaatttg tagtttattt atttttattt agtattttttt tttgtaaattt	2220
tgaggtttaga ttttttattt aaaagtgtat tttaaattttaa gtaataatata gtaaatttttt	2280
ttttgtaaaa gtttagtattt atattttaa ataagatata ttgaattttat tttagtgaattt	2340
atataaaagaa aataagtgtt aaaaattttaa ggtagtttag ttttttagttt ttttttaagat	2400
taaagagaag agattaaata tagtattttt gtattgaggt aaggtttttt gtgttagttta	2460
taqaaatttaq	2470

-continued

agtttatgtt atggtatata tggttttggg atttgttg ttaaaattga ggttaaaaaga 1200
tattnaaga gaatggatga tttaggagt agagattgt aaagagaaat gaagtagagt 1260
ttggtagtta ttatgatgg gaaagaagag gagagataaa gaagatataa aagatagtta 1320
ggtaagagga tttaggaag aattatagaa tgtaggagt tatattaaga ttaattaagt 1380
aagatttagg agatttatat tttttttta gtttaggtta aattatttg gaataaataa 1440
aattttgtt attttaatta ttaatagtt aaaaagtatt aagttagttg tatttaagta 1500
atatttaaat attttgatataaaaaatg ttttgtaat atgaaattta ttatataaaa 1560
taaggtagat aggttgataat attggtagt tacgataata ttggagttt gtaattggaa 1620
gattttatta aaggaaattha ggggatttt ttagattta gttagtttta taacggttag 1680
aattttatgt aaatttggta taatthaatg atatttgagg aggaagggga gttttgtagg 1740
tagggattat ttataaaagt tttggttga aaaaattga gtttigtgt ttaattttag 1800
gtttggttat gattnaattt ttgtttttt attttaattt attttataat ttgttaatga 1860
atatttataa ttgttttaa tataattt atgtatattt ttaggattaa aaaaaaaaaagg 1920
tatgttaaaa ataaaagtat tatgtaaatg taagttat tatgaaaata tatatgtat 1980
ataattttt tagtaagata tagggttt ttagttttaa gattnnntt ttgattaatg 2040
ggtaaggggt gaagaagtaa tgtagttaa ggagatattt aaaaataaa ggaaaattt 2100
ataggagtga ttattttttt gttttatatt tttaataaa gtaggttga aaatatttag 2160
tatattataa attatatgtat agaggttaggg atttgataga atttgaataa tgtgaatttt 2220
aaaattttt attttaataa aaattaatag gtaattttt attaaaataa taaaataaaa 2280
taagagaaaa ggttagtaata gaaaaaaa tgggtatgta taagtttatt gagatataga 2340
agttataattt ttatataattt ataaaaagag tcggatgggt aagatgatgta gagattgtta 2400
aaagtatTTT ttattatagg aataaaaaaa ttaattttt agatgatgta ttaattttt 2460
aattttttt 2470

```
<210> SEQ ID NO 7
<211> LENGTH: 2229
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 7

tttttttcg gcgttgggtg gtgcgggtt gggtaggtg gagaagtcgt ttttgttaa      60
ggtagataga cgtgttgggg gtgggggtcg gggtagggg cggtgtattt aggggggtcg      120
tgtttttttt tggatatagtt ggaagtttt ttctgtattt taaattttt ttattttttt      180
tgagggattt gtttttaggt agtacgtaa ttgttggttc gggtttattt cgtatttttt      240
tattgggtga ggaaggagta tttgaatgg agatgggggt gtttccgtt tatataattt      300
tagagaagag gtgtgtcggtt ttgttattttt ggaggtcgcg gtaattgtata tttagagaaga      360
tttcgggtgt agttgggaag gtttattggt tggaaagagg tgttttttt tttagttaaa      420
gggtttgtt tggaaagggtt gtttttattt tgtttagtgg tattatagga cggtcggttt      480
ttatcgaat ttttcggac ggtatttatta tatagtcggg ttttcgtagt gttggttttt      540
taatcgtat attgttattt cggtgaggat ttgtgttggat ggtcggagaa ttttcgttg      600
```

-continued

cgggcgtata	tggtttaggtg	gcgttggta	ggcgacgttc	gggtgttagga	cggcgttttt	660
atcgtttat	ttaaatcgt	tgttggtt	taggttttc	ggtttttga	ataggggttt	720
gggggttaa	ggacgtttag	gttccgggg	taggaagt	ttttggta	agcgttttt	780
tttttttcg	gtatataattt	ttttatttat	ttatccgtt	tatcccgg	gcgagaggtt	840
tattaaggta	gggcgcgtt	ttttatgaa	ttatctaag	gttttgagt	cgcgggggtt	900
tcggtaatt	atttttttt	tttttttgtt	ttaggtattt	tagtttaggg	gtttgttagag	960
aagttcgaag	tccggataaa	cgcgtcggac	gttaataatt	ttttattttt	ggtagtagta	1020
aaggtaata	tatTTTATT	ttttattttt	gtttgttatt	aaaataaaagt	tgcgcgcgtt	1080
tgaggtagg	aaggcggtga	gatcgagaag	aagggacgtt	tcggagaaag	tgcgtttagt	1140
tgatTTTGA	aatttagagtt	ttcgggatt	tcgtcgagat	ttttgttagg	gcgttttaat	1200
ttgtttttt	attgcgtgtc	ggcgtcgtag	cgcgtcggt	ttagggtttg	gtgatttcgg	1260
tttagttcgg	cggtcgcggc	gaggTTTTG	gcgttagtcgt	ttggaatttc	gtattagaat	1320
cgggatcgg	taaatgtttt	ggttgaagt	ttatTTTTT	taagaatat	tgttgttagg	1380
aataaaatgg	ggTTTTCGGT	gttcaagt	atTTTTGAA	atTTTTTAA	aataatttat	1440
aaaaaaatgtt	tttggTTTAA	cgtttataaa	cgtttaagga	aatatgtaaa	tggttggTTT	1500
ttttatcgag	atggcgtttt	taattaatag	tgtatataata	tataataattt	tttttaattt	1560
tttttttag	agtttaagtat	tttattatat	gtaaattata	ataaagaaaa	gattgtgtaa	1620
gattatgtaa	gtcgattgtat	ttaaaatattt	gagTTTTAAT	ttaggTTTTT	tgtttttta	1680
ttaataattt	tttgcgtttt	gattagattt	gtgaagttagg	ttatggaaat	taataaagta	1740
aaaaatttttt	agtatTTTTT	ttcgTTTTT	ttttttttaa	aattaaatata	tagtcgtttt	1800
ttttttagta	ggTTTTAGTT	tttaggttoga	gttttttgc	gattattta	tagttattta	1860
tagtagttgt	tgttgcgtttt	gtcgTTTTT	cgTTTTGTT	ttttttgggt	cgTTTTGT	1920
atataaaata	tatTTTTAGTT	ttttaattaa	atTTAAATAC	gatttcggta	gaatttatat	1980
atTCGTGgt	gtatggattt	tgcgggtgt	ggggaaataa	atTTTTTG	gtattttattt	2040
attgagtttta	attcgaaaaa	tcgggattgg	gttttttagc	ggtatTTTGT	gggttttaat	2100
ttggttcgg	tttttttaga	ttttggcggt	gagagcgtt	ttttgcggg	tgggtggacg	2160
gagaggtaat	aatttggTTT	taataaaaat	ttgtcgTTT	cgaatcgaaa	gcgaaaggga	2220
agggagaag						2229

```

<210> SEQ_ID NO 8
<211> LENGTH: 2229
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

```

<400> SEQUENCE: 8
ttttttttt tttttcgtt ttcgattcgg tggcgatagg tttttgttga aagttagattt
ttatTTTTc gtttatttat tcgtaaaagt agcgtttta gcgttaaggt ttggggaggc
gcgggttagg ttggagTTT tgggggtcg tttagggtt tagttcgat ttttcaatt
agatttagtg gttaaatattt agagggtattt tattttttt gtatcgatatttattatgtt
ttacgaaatg tgtaaatTTT gtcggggtcg tatttgaatt tagtttagaga attgggggtgt

```

-continued

```
<210> SEQ ID NO 9
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 9
```

-continued

ttttttgggtg agatatgtt tttataagtt ttaatggaga aaaatgtaa tatttttttt
tttgaattt ggttatttga gtaatgagaa aatagttatt ttttttagga tagtggttt
taattatgt tatgtgttt ttttagaaaa tttaaaaat atatatataat taatgtttt
gtgttattt tagggattt aagttttga atacgaattt tgtatttagta tttttaatt
attnaggtga ttgtgtatgt aaattatgtat tgagtttat ttttttaaga tgaataaaat
ttttttagt attgaattt aaaaattttaa ttattnaaat taattaaggg tatgggatt
aataaggtat agggaggtt ttatattata aaattttttt tttaaatttat agtttattgt
ttatnagttt tttgttattt tagaaaaggg tgaaaaataa gtaaaatttaa ttattnnag
tttggaaaat tatttagaaa tgaagatgac gatttgaaa tattgttaat attattttag
ttataaataa tgtttaataa tatttattt atattgtatg atattnnnn atatgaaat
tataattnaa aattaaggta ataatgtatt tagaatattt tatttattt tatgtattn
aagtaggtta gaaattaaga tatgagttt taagtatgag atgttaaggt gtggggtag
aaattatattt gtattnnattt attaataattt aatataattt ttaattnnattt atattnnaa
ttttattnn tataattnnaa attattnnattt attatgttta taaatataag tataatnatt
ttttagtatt tatttattt tatttattt tatttattt tataagggtt ttttttaat
tatttattt taaattnnattt tatttattt tatttattt aataagttt ggaggtaggt
attgttattt tttattnnattt ataaatgagg aaattgttta tagttataaa gttattgtgt
tagatnattt agaagtttaa tatnatttgg gtgaatnntt gtataaaaat agagagatag
atatgtataa tagtttattt ttatnattg taaaagttt taaatnntt tagaaattn
tttggaaaat ttagttaaaa atcgaggat tttttttt gttatnattt tataagggtt
attnnattt ttaataaagg atgaatnattt aatgtggat tttaaagggtt aattnnatt
ttttgaatt tttgatnattt ggatttggaa tttgttattt gtttttaaagt ttttttaagg
attnnatttga ttaatttaggt ttagaaattt tttggatttta ttgtcgaagt ttgagaattt
aagtttgggt tttatnccgg tttatnagaa agggttaatg aagtattatg gatagaattt
atnacgtttt agttagttt tttttttaga agttaatnagg tagtaatataa gttagaaattt
gtgattttagt ttttgggtt tgaagtttagg tagaatttta taggttttta gtgtgttt
tgacgagatt tttttttgg ggttaagttgt ttgtatnntt taaatnntt tttttttat
attnaaatgag ataaatnntt ttgttttata ggggtgtttt aaagattttaa taaaataat
atgttttattt ttatnagtttta taatgtttga tatttaagaa gtaaaggata tatttattt
ttattgttattt aatttagaaag tatnnaattt tgaaggagat aagatnntt gttgttagt
tatttattt ttttaggtt atttatttattt tttaaattt ttttgggttga gaataattt
taatnnttattt attnnatttga ttagtattt tatttattt aatgtatnntt tttttatnag
aaaaaaaat gtttttaagt tttttggaga aaatnattt gtatttattt tattgaaaaa
tttaataattt ggatttgggtt ttttgggtt aattnnattt tagtattttagt ataaatnnaag
tgttttaggtt taagaagattt gaaagtttagt atgtatnattt aagaatttgg
taatnattatg gtatnattttagt gttatnattt tagtaatnattt atgatacgtt atgttaaggt
tagttttagt attnnattt aattnnattt aatgtatnntt tttttatnag
gtttaaaggat ttaattnnattt gttatnattt gttttttggta ttttagtgtt
gtttaaaggat ttaattnnattt gttatnattt gttttttggta ttttagtgtt

-continued

agtttaaata atgtatataat ttttatttat atattttagta atataatttt ttgtttaatg	2340
agtgtatgtt ttttggattt tggtgggtt tgtagttt agaatttggtt ttttgggtt	2400
attataatata taagtataga gtaagtgtaa taaaattgtt gtatTTTt tgaaaagggtt	2460
tgtttttaaa ttgtttaata atttaaaggaa tttttgttga agtaatcgta ttgtttaatt	2520
agttataatt agtaatataat tttttggag ttttaattta tttttggtaa aacgttttag	2580
gaagagtata tattattaga aagtatgtt aaaaattttt tagtagaaaa tttaaaaata	2640
gtttttttt gttaagaggt tttttaaaat ttatTTTata tagttaaattt ttgaaatttt	2700
agtaggtttt gtTTTatttat tataattttt gtataaaatat ttTTTaaaggat ttgtttttt	2760
gttttaagta tgatTTTttt ttataagttt gattagtat tatattagtt ttgtttagga	2820
aatatgatatg tttttatTTt ttgtttaga gttgttaaat ttgttattat attttatgtt	2880
tttttttgt tgaaagttt tagcgaaaga aatttttaat tttttgtttt gtaatattag	2940
ttggtagttt tatttaatgg gtatTTTgtt tttttaaaaga atttagttgt ttgtttaga	3000
agtcgatttt ttgtatTTTt taacgtttgg ttttaattgtat ttgttttaat ggagtttgc	3060
tcgggtgagga gcgagatgtt atcgattaga atgttggat ttgtttagttt attgttagga	3120
gtgagagata ttgagattta gaaatTTTt gaggtggag gggagaggga tagttcgg	3180
cggaggcggaa gatgtaaat aaagggtatgg atTTTatata gaaaaaaa aaagatttcg	3240
ttgaggtatt gaggtgtt acgattataat tttttaaagg agaagttaaa aagtaaggaa	3300
gtggggaggag gttggagggtt aaagtattta aaaggattat tcgggtataa ttgtttttt	3360
tgttgggtt tgtaaaggat agatagtttc gtttttaaag tatatgaatg ttttttttaa	3420
gtgatTTGGA atggatatta attgtttgtt aaatgttattt aaatgttttt ttAAatttag	3480
gggatataga aagaggggtt taaaaggaga atttaaatag aaaaaggggag gattcggagg	3540
ttttgaaag cggggggaga agaaggagga gggataatag agaggaatag agaaggagag	3600
cggagagaag ataaataaaa ataaaatag gaattattga ataattataat attaaaaaga	3660
aagttttttt ttatgggtt tttaaaatat tgagattgtt atagtgtt cgggtatgg	3720
agaaagatgt tttttttat tttgttttc gaaagtTTTt ggTTTcgta ttggcgatTA	3780
aaattttattt aggttaaaga gtgtgttttta ttgtttaga aatgttagtag acggaaaggcg	3840
ggTTTcgta tgcgtttgt tttttcgTTt ggagagaatg aaagaaacgc gtagagtt	3900
agatTTTgtt cgagttatgt ttttttcgt cgTTTtaggt tATCGGTat tcggtaaaga	3960
ttcgagtaag gaacgttaggg ttattgtttg ggttaataaa tggagttcgt ttttttttt	4020
tcggacgtcg ttgttccgtc gatgtttcg gtaattttt cgcggcgat gtagaggagt	4080
tttttttttt ttttttagatt atttgtttcg attaatttga ttttttaat atatttgc	4140
gtatTTTTta ggtggatata ttaataggtt acgggttggaa gaggagcccc tgatgaggag	4200
agggatttaa atttgcgaac gtttgggtt ggtcggagtt gcgggggggtt tgggaggaga	4260
gagggggagaa gagagaagga aggagagcggt ttgtcggat ggttgagttt tttcggcgag	4320
tagttttggg gttgtacgtt tttgtggag atgttggat ttttttaggt tcggtaagag	4380
cggTTTtaat attatcgTTt ttatTTTTt tttttgtaaa ttttttagaga aacgttttg	4440
gtttttcgat cgcgatattt ttagttgttata ttttttata gtttagggcg cgcgtttcg	4500
tacgttggag cgtcggtcgat tagtaggacg tttttttcg cgtcgatcg ttttttttg	4560

-continued

ttttgtgtt gttgttttt tgatatttc gttttatta ttttagttc ggagagacgt	4620
tattnagtcg cggttcgtat tcgcggttcg gggtaacgcg cggaaaggagg gcgttagttc	4680
ggatttcgtt ttcggtaggg ggcgtttgg agcggagagt gaggcgaatg gtatatgagt	4740
gtgcgggtag ttatatttga agttcgttgcgatt ttttatttga gttatgttgc gtttagttt	4800
attcgggtta gcgttggcg agcgagtttta tttgtggtt tcgcggcgtt ttttttttgc	4860
tattnnntgta ttatattcgtc gattnnnntt tttcgggtt tttatgttgcgat tttattaaatt	4920
agagttcgtat tgtnnnntt tacgtgattt cgggcgggtt gaggattgt tgtnnnntaa	4980
acgttagagg gatgcggcg gtagagttcg agaggcggtt gtcgggttgc gggcggttt	5040
gattnnnntt ttattnnntt tttcgggtt ttattnnntt gttttggat tttcgttttt	5100
ttttgtttttt cggtnnnnttta gagtnnnntt tttatgttagt tagtnnnntc cgtnnnntcgc	5160
gtagtnnnntt agcggacgtat ttttcgtttt cgggggtttagt tttattnnntt ggtatgttgc	5220
gaaattnnntc agattatgcg cgggtttggat tttattnnntt tcgtcgggttgc ttattnnnttgc	5280
cgtatgttgcgtt tttgtgtcg tcttcgtggg gatgttttagt agttcgttgc tccgttttcgc	5340
cgattnnntgtt ttttcggaa gtcgtttgtt gttgttagt tttatgttagt agttatgttagt	5400
ttgtgggagt tttcgcggta gttgttagt tagttagtgcgat tttattnnntt gcgagggttgc	5460
ttgtgtgttgcgtt atttatgttgcgtt gttcgttgcgtt tttattnnntt gttatgttgc	5520
attnnnnttaa gtgattgtta aacgtttatc ggttggaaatt gtttggtaa gtttggaaatt	5580
ttcgttttcgtt attnnnnttaat ttcgttagaaag aatacgcgtt tttatgttagt attatgttgc	5640
tttagcgcgtt tttattnnntt tttattnnntt gattnnnttaat attattnnnttgc	5700
tttattttaga gaaataaggg gaattgttgcgtt aggttcgggg gttgttagtgc gtttggat	5760
gggttagaaag tttatgttagtgcgtt gtaggtttttt tttatgttagtgc gtttggat	5820
gaggattnnntt agaaatatttta aacgggttgcgtt tttattnnnttgcgtt tttattnnnttgc	5880
taattnnnttgcgtt gtaggtttttt gttatgttagtgcgtt tttattnnnttgcgtt tttattnnnttgc	5940
aggatgagga agaaataatttta ttatgttagtgcgtt gtaggtttttt gttatgttagtgcgtt	6000
tgtatgttgcgtt gtaggtttttt gttatgttagtgcgtt tttattnnnttgcgtt tttattnnnttgc	6060
ttgtatgttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6120
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6180
ttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6240
ttttttttttt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6300
ttttttttttt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6360
ggcggaaatcg atttatgtgg gctgttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6420
gtttaggtat tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6480
gtttaggtat tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6540
tttagttagtgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6600
tttttgtaaaatcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6660
atgtatataa gttgttagtgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6720
gggtgaaatcg atttatgtgg gctgttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6780
atgtatttattttaatcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgcgtt tttattnnnttgc	6840

-continued

tgggtaata gattgtttt tttttggta gttttaaat ttggtatTTT aataaaattt	6900
aatatgtttt tataattttt tgatttatgc gstatatgtgt gttgttttg aaagaataag	6960
tttattttg ttattgtta attattttt agatgtttt ttatggtaat aattatgagt	7020
ttgtaaaaat aatttttggaa aatgtttagt gttttgtagt ttaatataga ttggttgtt	7080
ttatTTTtag ttttgtatt gttttaggaa ataattaatt taaatgtgaa gttgatattt	7140
gtaattaaga aattatataat ttatttagata ttttaaaggg gattgtataa attaaagaga	7200
ataaatttgtt tttgttagata gggtgttaag aatttggtat ttcgtttta ttttgttaa	7260
tttagaggtg attaattttt attttagttt aatagattat tatagaaaaat attgtgtttg	7320
tttattttta ttatttgaggt tttgttttt tttgtttgg atatattta aataaggggt	7380
tgttttagtc gttgaagtaa aagaataatt aaagatgggg aaatgttaaa agggtattta	7440
gagattatta ttagttttt tttaaaatgt ggagttttgt gggtataaat attgtttatt	7500
taatgagtaa aaaataaaaaa taaaaaaaaa ataggaagta aatgttaagt ttttatttt	7560
tattgttagt attaacgtaa gttttaaaaa atagtattat tagaaaagga tattaaagga	7620
gaattgatta gaaaagaatt gtggaaaatg gaaacgaata ttgattttt aatttagattt	7680
tgaggttatt agtagatagt gattttgttag tatagttata gttgttgat ttaaaaattta	7740
ggataagtat tttaaagttt taaaatgtg tttttttt ttaaaaattt gtaagatgtt	7800
ttaatgattt gagtgtttt tttgaatttg agg	7833

<210> SEQ ID NO 10
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 10	
ttttaaattt aaagagaata ttttagttat taaaatattt tatagtttt taataaaaaa	60
aagtattattt ttgaagtttt aaaatattt ttttaaattt taaatttaat aattatagtt	120
gtattgttaag gttattgttt attgataatt ttaaaaattt gttaagtgtat taatattcgt	180
ttttattttt tataattttt ttttagttaa ttttttttta gtattttttt ttgatagtgt	240
tatttttaa agtttgcgtt aatattgata gtggtaatg aaagttaat atttgtttt	300
tgtttttttt ttatTTTtat ttttggtaa ttaggtggat aatattttagt attataaaat	360
tttatattttt ggaaaagagt tagtgtatgat ttttgaatat ttttttatta tttttttatt	420
ttaattgtt tttttgtttt aacgattgaa ataattttt atttgaatg tatttagata	480
aagaggaaat aaagttttaa taataaaagat aaataggtat agtgttttt gtgtatggtt	540
gtttggttta aatgaagatt gattattttt aagttaatag ggggtggaaagc ggggtgttaa	600
gtttttgata atttatttgtt aaaatttagtt tattttttt agtttatgtt gtttttttaa	660
aaatattttgg taaatatgtt aattttttgtat tgaatgtt aattttatattt ttaagttagt	720
tattttttaa aataatgtaa gggtaggaa tgaagtaat tagttgtgt tggattataa	780
agtttataat atttttaaaa attgtttttt taggtttata attattattta taataaaagta	840
ttaaaaaagt gattaggtaa tagtaaagtg aaattttattt ttttaaaaat aatataatgt	900
tacgtatgaa ttaagaagtt atagaaatattt gttgagttt attaaaatgt taaatttaga	960

-continued

aattgttaaa aaagagaata atttatttat ttaaaatcaa tagggttgta tattttatt	1020
tgttttgta aaggataaat tagaatgatg tataataatt ttttttttgg tatttatatt	1080
agtaataatt aggaattata taggtttta ttttgagtta tagttggta tttttttt	1140
ttaaaagtta tatataatttt agtttatata tattttgaa agatattta tttagagttta	1200
gatthaatta tagtaaaatt atatttata aagatgaaa attatataata ttttatatta	1260
taggttgtta aatcgaatgt tatgttagt aggagtgtag taattttat tttttggttt	1320
tatthaatta ggaagttta gtagagcgaa gtttgttaag cgttcgtcgt tagaattga	1380
aggaattcga gcgagtaaga agagtgttg atttattta tagaagtgg tttagaatgt	1440
gaggagttag cgtttattga agtcggttc gtttccggtt cgtttatatg gagtttgatt	1500
agtttagtt atgtttattt cggtttggaa gattcgtaaa gtgtttttt ttttaatttt	1560
tttgtattat tttgaagttt agggaaagtaa agagaggggt atatttggat tgtaaaattha	1620
atgttttttgc tcgtttagga gagaaggaa tgagagagag agagatagat atagatagag	1680
agagagagag agagagagag agagagagag agagagagag agaaattta ttgaaattta	1740
gttttttag aatttgcgtg atttgggtt taacgggaga tttagtgcgt tttatggat	1800
ttttgttagg aatttagcgt tttttgttag ttattatttgcgtt atttatttgcgtt tttcgttta	1860
ttttttttta taaagttatt tttttttat ttttagtaaga tttttttttt taatgtatgt	1920
aaagtttttgcgtt ttttaggtt ttttttagga ttgggtttt tttaaaatgt tgaatttgcgtt	1980
aaattatttc gtttaatatt tttaaaattt ttgcgttagtt taatgtaaac gtaagtatgt	2040
aaaggtttttgcgtt ttttaggtt ttttttagga ttgggtttt tttaaaatgt tgaatttgcgtt	2100
tttgcgtatgt ttttttttttgcgtt ttttttttttgcgtt aatttttttgcgtt ttttttttttgcgtt	2160
gttaggaggtg aggggttgag gaggcgctt agggtaggtt gggttgcgtt ggatacgctt	2220
gtttttttgc gggatcaaag ggtcggggac ggggggttttgcgtt ttttttttttgcgtt	2280
gtcgggtggc gtttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2340
gttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2400
ttagttgttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2460
agtagtaaac gtttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2520
attttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2580
gttagttaat ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2640
tcggagcgag ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2700
aaggagggag ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2760
tggagttcgat ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2820
tttcgagtttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2880
tcgggggttac gtttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	2940
ggagggaggg gtttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	3000
agatgggttcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	3060
agaagttcgat ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	3120
gttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	3180
tttcgggtcgat ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt ttttttttttgcgtt	3240

-continued

-continued

tatgtaaata	aaaatatata	tattattn	atttgttatt	aggtattaaa	gaagtagata	5580
agattgtatt	agtaattgga	ttagtgttt	aattttttt	tagtaaggta	aaatttagtt	5640
atttattaga	attaaattta	agtttatgaa	ttgtatttt	tattgcgtat	tatatgattg	5700
ttagtaatat	gatataatta	tattatgtat	ttgtaaaatt	tttattttaa	aatattatat	5760
tatattttt	ttaattttt	ttgagttaga	atattttt	tgtggatat	atattttaga	5820
attgatgtag	aggagtagag	tttagttgt	agatttttta	gtagaaatag	tgtagatata	5880
tttttttag	aaaatttaag	aatattttt	tttttatgg	aaagaatatt	attataaagt	5940
gtgagattat	tttagttta	agtaggggt	ttgggagta	tttttaata	agaatagttt	6000
aagataaaata	aatgaatttg	ggaaaataag	atattttgtt	aattagaatt	tttattttt	6060
ttatgatttt	atatttttg	attgttttaa	taaaggtaag	atgtattttt	tgttttttag	6120
gtgttaggt	ttgtgtttag	taggatagaa	tatgttattt	ttatthaatt	tttaaaatat	6180
ttttatgaga	taaagaatata	tattttattt	tatataaaag	gaatatgtt	ttgaaagtat	6240
taggttaattt	gttttaagaa	ataaatttcg	ttagtgatat	tgtgggatt	ttgtgaaatt	6300
ttgtttgatt	tttagtgtata	agatataagt	tattaattt	tgtgttattg	ttgtttgtta	6360
gttttgaga	ggggaaattta	attgggaacg	tattagttt	gtttatgata	ttttatttgt	6420
ttttttgtg	gagtcgtagt	aaggtttaaa	tttaatttt	taaatttcgg	taataagatt	6480
tagtgatttt	tgaatttggt	tgattatatg	aatttttgta	gaaattttga	aataatagat	6540
aaatttttaag	tttttatttt	agggatttag	aaaatttgg	gttggggttt	gggatttata	6600
tttaatattt	tattttggtt	ggagaagtaa	ggtattttt	atatttat	gataaaatgaa	6660
aggatatttc	gattttgtt	tagttttat	agagaggtt	ttgagatagg	ttaaaagttt	6720
ttattttagt	taaagatgag	ttgttgata	tgtttgttt	tttgtttta	tgtataatgtt	6780
tattaaatat	gtattaagtt	ttaaatatgt	ttgatatagt	aattttgtga	ttgttagataa	6840
tttttttatt	tgtaaaatgt	gagtaataat	aatatttgg	ttttggggtt	tttttaaaga	6900
ttaaaataaa	aatgtatgg	ttaatggtag	tggatttgg	gggattttt	atataaataa	6960
gtgaatggag	gtatgaataa	ataaatata	aaagatgtgt	gtatttat	ttatataat	7020
aattaaaaat	agttaaagat	gtataaatta	aagttaaatg	tatgtgatat	tgaagtat	7080
gttgattatt	gataatgaag	tatgtatata	tttttaattt	tatattttaa	tattttat	7140
ttaataat	atattttat	tttagttta	tttaagatat	atagatata	atagaatgtt	7200
ttaaatgtat	tatttttta	tttttaatgt	ataatattt	tatgaaaaag	tatttttag	7260
tgtgttagtaa	atgttattaga	atattattt	taggttaat	gatattgata	atgttttaga	7320
gtcgttattt	ttattttgg	ataattttt	aaattgagag	taattaaatt	tgttatttt	7380
ttatttttt	ttataatgtt	aaataatata	taaataatga	gttgtgattt	aaagaaataa	7440
ttttataatg	taaaagttt	tttatgtttt	attgattttt	atgtttttaa	ttaattttgg	7500
tagtttaagt	ttgtgatttt	agtgtttagg	aaagttttt	ttattttaga	gtagtgggt	7560
ttagttatga	ttttatatta	taattattt	gataattaaa	gaatattgt	gtagagttcg	7620
tattnaaaga	tttggaaattt	ttagaagtga	tatagaagta	ttggtatata	tatattttta	7680
aagttttttt	ggagaaatata	atagttatga	ttgagaatata	ttgttttggg	gaaagtgatt	7740
atttttttat	tattnaaata	gttaagtttt	aggagtaaa	atatttat	tttttttat	7800

-continued

-continued

tcgcgttagta tcggaggaag aaagaggagg ggttgggtgg ttattagagg gtggggcgga	1980
tcgcgtgcgt tcggcggtt cgagaggggg gagagttagt agcggggcgcc ggggagtagt	2040
atggagtcgg cggcgccccg tagtatggg ttttcgggtt atgggttgt tacggtcgc	2100
gttcggggc gggtagagga ggtgcggcg ttgttgggg cggggggcggtt gtttaacgta	2160
tcaaatagtt acggtcggag gtcgatttag gtgggttagag gttttgtacg gggagtaggg	2220
gatggcgccc gatTTTggag gacgaagttt gtagggaaat tggaaattagg tagcgttgc	2280
atTTTcgga aaaaggggg gTTTTTggg gagTTTTtag aaggggtttga taattataga	2340
tTTTTTtg gcgacgtttt gggggTTTgg gaagtttaagg aagaggaatg aggagttacg	2400
cgcgtataga ttttcgaat gttgagaaga tttgaagggg ggaatatatt tttttagat	2460
ggaagtatgt tttttagat atataaaatt tacgaacgtt tgggataaaa agggagtttt	2520
aaagaaatgt aagatgtgtt gggattattt agtttttaat ttatagatat ttggatggag	2580
tttattttt ttatttaggg ggattatttag tggaaatttg tgggtgtatgt tggaaataat	2640
atcgaatata aattttgtatc gaaattattt agaagcggtc gggcgccgtt ttttgcgtt	2700
tgtatTTTT ttatTTTggg agattaaggc gggggggattt atttgaggtc gggagttcga	2760
gattagttt gttaataggt gaaatttcgt ttttattaaa aatataaaaa gtagtcgggg	2820
gtgggtgttag gcgtttgtaa ttttagttt tggggaggtt gaggttaggg aatcggttga	2880
atcggggagg ttgagggtgt agtgaatagc gagatggagt tttttattt tagtttgggt	2940
gatagagtga gattttgtcg aaagaaagaa agagagaaaag agagagagaa aaattattta	3000
gaagtaatta tataattgtgt ttatTTTaa ttgagtaggg taaataaata tatgtttgtt	3060
gttaggaattt aggaataat gagttatatt tatgtgttta tttagaggtt aatatgtgt	3120
tattttttt ggaatatttg ttaatatttt tgTTTTTTt ttatTTTtag tttatttgat	3180
atagttttt tgtgataaga gtttttaattt ttttattttt gaatagaggtt gttttttttt	3240
ttttttttt tgTTTTGTGA gggagttagg ggaggattta aaagtaatta atatatgggt	3300
aatttagtat ttttaaaattt ttgttaatag tttgaattcg ggagtttgggtt ttgttagtt	3360
tataatattt tagaagagat tttatttgat taaaataaa aaggaaaaag aaaagtggat	3420
agttttgtata attttaatg gagaaggagg aagaatatgt agaaaagggg aaatgtatgtt	3480
ggtttagaat tttaattata ttgggtgtta atataggaat atttatttat ataataatttt	3540
aaagtattaa atttatattt gtatattttt aatggatattt atttattttt gggtttaagt	3600
attttatata tttaattttt attgattttt tttttttttt ttttggattt ttattatgtat	3660
ttaaatattt atatatgggt ttttttttagt tttttttata ttatgaaata taagaaaaat	3720
tttaagggtt agttttatgt ttaagacgaa ggattttattt gaatatataa aataataat	3780
atattgtat attttgggtt tttttttgtt gttgtatattt gggttgggtt tttttttttt	3840
ttgtttttt gaaaattttagt ttagttttt ttttttttaga taggatttaa taattataat	3900
ataattttatgt ataattttttt gatttaggtt aattatgtaa tttgtgttta gtatgaaatg	3960
tatTTTTAA taagtaattt ttttttaata ttattttttt taaattaata taataaataa	4020
tagttttttt aaaataaaattt gtttattttt attatgtatgtt atttaatattt taaggttgg	4080
atgattgtat atagtattt aaaattttttt ttggaaatg gttttgtttt taagatgttt	4140
taggaattaa agaggtgattt atttttgtt taatgaattt ttaaaattata aattttggaa	4200

-continued

gtgttttagt tttttattgt tggtgttata aattattata aatgtgttag ttaaaataaa	4260
tataaaaatta ttatTTTata gtttagaga tttagaagtta aaaatgggtt tataaggTTT	4320
tatTTTTTTT ggaaattttt aggggtaatt tgTTTTTTg tttTTTTtag ttttagtga	4380
ttatTTAAATT ttTtggttta tggtTTTTgt atTTTTTTg tggtTTgtgt ttTTTTTTT	4440
gtatTTTTT tttgattgtg atTTTTtaat aaaaatattt ggggttatgt tgggtttatt	4500
ttgaaaattt tggataattt ttTTtaagat tattaattaa attataattt taaagTTTTT	4560
tttgTTTATAT aagttaatgt attaaaagtt tttgaggatt aggatataga tattgggggt	4620
gggggggtat tatttagTTT attatagaa ggaatttttag ggttaattaa attagTTTTT	4680
ttatTTTata tttgaagaaa ttgaagTTT ggaattggag agtattatgt taaatgaaat	4740
aagttaataa tagaaaagata aatattatat gtttttattt atttgtgaaa tataaaaataa	4800
ttatTTTTT agtagtaaag agtagaaatgg tggttattag agttggggg tggaggaat	4860
ggggagatgg taattaagat ataaagTTT agttaagatg ggaggaataa gtttgattgt	4920
ttttttgag atgtgtttta tagtatgatg aatatagtt aatagtaaat tttaatgtt	4980
tttatttgc aaaaatgtt aatatttgc atgatggata ggttatttag tttgatttaa	5040
taatTTTTta ttgtgtttaa agattataat ttTatTTTgt attatataaa tatatataat	5100
tgtattttt taatataataa ttTTaaaatt aatataatgt aaaaagaaatt gaagtttaat	5160
atTTTTtagaa gttaagtgtt aTTTTaaagt tttgtgagaa ttgttttaa taaataaaata	5220
agTTTTTTT tttaataat tattatattt tgcgTTTgga tatatagttag tgaataaaaa	5280
aaaaaaaaaaa aaaaaaaaaatt tttaggttta atataattt aggaagaaat tttagtagtt	5340
gtatTTTTtagg ggaaatatag gaagtttagt tggagtaaaa gttagTTTgt ttttggTTT	5400
ttgttattttt gttcgtgtt tatagtgtt ttgtTTTgt acgatagTTT cgtagaagtt	5460
cggaggatAT aatggaaattt attgtgtatt gaagaatggta tagagaattt aagaaggaaa	5520
ttggaaatttG gaagtaaatg tagggtaat tagatattt gggTTTgtgt gggggTTTgt	5580
ttggcggTGA gggggTTTta tataagTTT ttttgcgtta tgTcgTTTt tattttggTT	5640
ttgattttttt tgTTTTTTT ggttagg	5666

```
<210> SEQ ID NO 12
<211> LENGTH: 5666
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

<400> SEQUENCE: 12

tttgttagag agaatagaat ggtagagtt agggtgggg tcggtatgac gggaaaggaa	60
tttgtgtaga gttttttat cgtaatgt attttatata aagttttagg tgtaatttt	120
tttttatatt tgtttttagt tttaatttt ttttttagt tttttatata ttttttagta	180
tataatgaat ttattatata tttcgaatt ttgcggagt tgctgtata ggttagaggt	240
atgtgaggt acgggtaaaa tagtaaaggg gttagggatag attgatttt atttaggtt	300
aatttttgtt atttttttt agatataattt attgaaattt tttttgaaa ttatgttagg	360
tttggagatt tttttttttt ttttttttt tgtttattgt tgatattta agcgtagaat	420
gtggtaattt taaaaaagag aaaatttggtt tgtttggtaa aataaatttt tataaaaattt	480

-continued

-continued

tttattataa	ttttagtttt	tcgggttaaa	gcgatttttt	tgtttaatt	tttcgagtag	2820
ttgggattat	aggcgtttgt	tatttttc	ggttattttt	tgtattttta	gtagaggcga	2880
ggtttattt	gttggtttagg	ttggtttcga	atttcgatt	ttaggtgatt	ttttcgttt	2940
tgattttta	aagtgaaggg	attataaggc	gtgaggtatc	gcgttcggtc	gttttgaat	3000
aatttcgatt	aaaatttata	ttcgatattt	attttaatat	atattataga	tttttattga	3060
taattttttt	tagtaagaaa	gataagttt	attttaggtat	ttgtgaattt	gaggtaagt	3120
agtttttagta	tatTTTtat	tttttaaga	ttttttttt	atTTTaaacg	ttcgtaaatt	3180
ttgttatttga	taaagagtat	atTTTattt	aatataaa	tgtttttttt	tttagatttt	3240
ttagtattt	gagagattt	tacgcgcgtg	gttttttatt	ttttttttt	gttttttaa	3300
gttttttaggg	cgtcggttagg	aggaggttt	tgattataaa	tttttttga	aaattttta	3360
ggaagttttt	tttttttcg	gagaatcgaa	gcgttattt	atTTTaaattt	ttttgtaaat	3420
ttcgTTTTT	agagtcgttc	gttattttt	gttttgcgtg	tagattttt	atTTTatttgg	3480
atcggtttt	gatcgtaatt	attcggtgcg	ttggtagcg	tttgcgttt	tagtagcggt	3540
cgtatttttt	ttattcgatt	tcgggtcg	gtcgtggta	gttagttgt	cgaaggtttt	3600
atgttgtttt	tcgtcgctgg	ttttatgttg	ttttcgctcg	ttcgttgc	ttttttttt	3660
ttttcgttagt	cgtcgagcg	acgcggttc	ttttattttt	tggtagttt	ttagttttt	3720
ttttttttt	tttcgggtgtt	ggcggaaagag	ttttttcga	ttttgtttt	taaatttttt	3780
ggagggatcg	cggtatTTTT	ttaggtttagg	ggacgtcg	acgcgtgtt	oggaggaggt	3840
gttattaatt	tcgagtattt	acgcgtatgt	gtattttga	agtcgttttta	ggttgggttt	3900
ttttcgggggg	tattagtcgg	aagtgtttt	cgtttagagtt	acgcgttgc	aggaaggagg	3960
atgggtttt	tttttatttgc	cgtttttgc	ttttttgtt	tttagtcgc	4020	
tttttcgtt	tgttagtaaa	ggcgtgtttt	agtgcgttta	ttttgtttaaa	aagaatttcg	4080
ttttcgttcc	gtttttttt	ttcgcgat	atTTTttta	attgttaat	tgaatcg	4140
tgtttgggt	tataggaaa	gtatggttt	tttttttaat	tataaaaaaa	agtaaaattt	4200
tttttttttta	gttgcgttgc	ttttatecg	aatcgaaattt	attgtacga	tttagaaagt	4260
ttttttattt	ttttaatttgc	ggagcgcggg	gttttatttgc	ttagaaattt	4320	
tagttaaag	gattttttt	ggagagtcgg	attgtttttt	ttttttttt	ttttttttt	4380
tttgcgtgt	aaacgggtt	ttgggttaa	gtttttttt	acgtgtat	tgtttggat	4440
aagagtagat	tttgaaaaga	ttaggttat	ttaatacg	cgggggagaa	ttttgtttgt	4500
aggtagatag	gaaaatgggg	agggagttt	tggaaaggacg	gattttttt	ttaaagttat	4560
aatttttaga	ttagaaaaag	tgttagtgc	ttttagtgc	gagttgtata	gtgatttaaa	4620
gatttagttt	aaatattgtt	ttgtttttt	tatattttt	atTTTttttt	tttttattga	4680
aaatattttg	tattttcgt	aattataag	ggggaaaggga	atatgagtgt	tttttgc	4740
ataggggtt	ttgtgagtt	aatatgtat	ttaatata	taagtttaa	aatagtgtt	4800
atataattta	agttaatatt	tgttagttt	tgaattatc	gttttgc	ttgggttgc	4860
atTTTtttt	gaggtataga	aagaaaatgt	tttggagtag	gacgcgg	tttataatttgc	4920
taatttttagt	atTTTggaa	gtcgaggcgg	gtagattatt	ttaggtttag	agttcgaggt	4980
tagttggtt	aaaatggta	tatttcgtt	tttattaaaa	tataaaaatt	agttggttat	5040

-continued

```
<210> SEQ ID NO 13
<211> LENGTH: 2470
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

<400> SEQUENCE: 13

aaagatgatt aaaagttaa ttgttattt gaagagttga ttttttattt tttgtataaa 60
agggtatTTT tagtagttt tgTTTattt gTTTattgg tttttttgt ggTTTgtAA 120
ggTTataatt ttTgtTTT agtaaattt gTatgtta tttttttt tgTTatttt 180
ttttttta ttttggTTA ttatTTgt gtAAAattt ttgttaattt tatttGAAat 240
gagaaaattt aaggTTata ttatTTaat ttgttagat ttttattttt gttatATgt 300
ttataatgtg ttgggtattt ttAGATTGT ttatTTAAA gatgtAAAat aaATAatGA 360
ttatTTTGT ggatTTTTT ttatTTTG agatgtTTT ttggTTgtA ttatTTTT 420
atTTTTgtT tattgTTAG aggaggGGTT ttaattatGG gtGAattTTA tatttattG 480
aagaggTTat gttatATgtA tattttata atataatttA tatttataAtA gtatTTTAt 540
tttttagtata ttTTTTTA ttaattttAA taatatttA gtaagttagt ttGAAGTAGA 600
ttgtAAgtgt ttatttataA attgtGAat gaattAAAat gaaagggtAA agattAAatt 660
atgatttaggt ttgAAatttA tatataAGAt ttaattttt ttaatttAAAG atTTTTgtA 720
gtgatTTTg ttTgttaggat ttTTTTTT ttttagatgt tattggatttG tatttaggtt 780
attgttagatt tttagTTgtG tagAAatttA tagatTTAAG atgagtTTT tgatTTTT 840
tggtagagtt tttaattgt tgaattttAA tattgttgTg attagttgt gttataattt 900
gtttgtTTA ttttggTgAA tggattttt attatAGAGG tattttttA atgttaAGat 960
gtttaAGAt tggatTTAAGt taaatttAtt aatattttt agttatTAAG taatttAAGat 1020
aggttaggatt ttatttggTT taaaatgtt tgatTTAAAt taaaAGAGA atgtggattt 1080
tttgaattttt atttggTTAA tttaatAtA atttttagtA ttttataatt ttttttAAAG 1140
tttttttattt tggTTatttt ttgtatTTT tttgtTTTT tttttttttt ttgttataAA 1200
taatTgttag attttggTTT attttttttt gatagtTTT atttttAAGG ttatttattt 1260
tttttaggtA ttTTTggTT tttagTTGAG tatagttagat tttaAGAtA tatagtgttA 1320

-continued

agtataggtt atttagtta atttttgaa taaaatgtat tgaatttat gtttagtaatt	1380
tttattttt atttttttat taaaagggtt taaagttttt atttatgtt tttttttatg	1440
tttattttgt taaaatgattt ttttttaatg atattttaga atttttagaa tattttatta	1500
tggaggatgt gtaagattag ttttttatta aataaaaaagt gtgaaaatgga atatgtat	1560
ttattaattt attttggtt taaaattttg tgattattag ataaaattta gaaataaaaat	1620
agtatttattt atataaataa attttttatta taatttatatt ttttaagttt tggtttaag	1680
aatgggtaaa atattttaa aattttgaag aaatttattat ttgatagaaa gtttaattta	1740
tttgcgaa ggttaatgtt ttttagatata attaaagttt ttttttttat tttatattta	1800
tttattttga attaagattt tattgttta tttttttaga tggttatttatt tgaataat	1860
tgttttgcgaa ttaaaaatttta gtatattaat ataattttttt tttaaatgttt taaggtttt	1920
gtttttttta tttttttttt taaaataaag tagttattaa attttttagt agtgaatttt	1980
aaaattttttt ttaattttat agggttaagg gtagttaagg atgggtttag ttttatatgt	2040
ttagttgtta aagtaagttg aggttattgaa gatggagaat tttaaattttt gataagagtt	2100
agaagataat tttaattttt ttataaaattt gggaaattttag gtagttaataat tgaaggttt	2160
aagatttgta ttttaatttg tagtttattt atttttattt agtattttttt tttgtaaattt	2220
ttaggttgcgaa tttttttttt aaaagggtat tttaaattttaa gtaataatataat gtaaaattttt	2280
ttttgtaaaa gtttagtattt atatttttaa ataagatata ttgaatttat ttagtgaatt	2340
atataaagaa aataagtgtt aatattttaaat ggtagtttag ttttttagttt ttttttaagat	2400
taaagagaag agattaaataa tagtattattt gtattgaggt aagggtttttt gtgttagttt	2460
tagaaatttag	2470

<210> SEQ ID NO 14
<211> LENGTH: 2470
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 14

ttagtttta tgaattatata agaaaatttt gttttagtat agtgatgtta tatttggttt 60
ttttttttta attttaaaaa gaattaagaa ttaattttagt attggagttt tatattttatt 120
ttttttatata gatttattga atgaatttaa tatattttat taaaaaatat aaatgttaat 180
ttttgttaaga aagagtttat atattattgt ttaatttaaa atatattttt aagtaaagtg 240
tttttattta agtttataaag aggaaatattt gaataaaaaat ggataaaatta taattaaaag 300
ttatagtttt gatattttta tatttagatgt ttttagttttt agttttgtaa gatgattgga 360
attattttttt agtttttgtt gaagatttga gttttttattt tttagtgttt taatttgttt 420
taataatttga ttatatgaag ttgttagttat ttttggttat ttttggattt ataaggttaa 480
aaaggattttt gaaatttttattt attaaaaaat tttagtggttg ttgtttttta aagaaaggggg 540
taaaggaaat aaaattttta agatgtttaa gaagaattgt gttaatatgt tagttttgg 600
ttttaaaata atattgttta agtagtagta tttaaagagga tgaaatagtg gagttttagt 660
ttaagataaa tgaaattaaa atagaagaga gaattttagt tttgtttgaa tatatttgtt 720
tttttataga tggatttaaaat tttttattaa gtaataattt ttttaagggtt ttaaagatata 780

-continued

ttttattttt tttataggta aaattttagga aatataatta tgataaaaaat ttattttat	840
tagtaatattt attttatttt tgaattttat ttgatagtta tagaatttta gagttagaat	900
ggattaatga gattatatat tttattttat attttttatt tgataaaaagg ttaattttat	960
atatttttta tggtgaaata gtttgaagt tttaagatgt tattaaaagg taattttta	1020
ataaaatgga tatgaaggag agtatttaat gaagattttt agttttttg ataggaagat	1080
ggttaataaag aattattaat ataaagtta attatattt tttaaaaggt tgattataat	1140
agtttatgtt atggtatatg tgggtttggg atttgggtg tttaaattga ggttaaaaaga	1200
tattnaaaga gaatggatga ttttaggat agagattgtt aaagagaaaat gaagtagagt	1260
ttggtagtta ttatgattgg gaaagaagag gagagataaa gaagatataa aagatagtta	1320
ggttaagagga ttttaggaag aattatagaa tggtaggat tatattaaga ttaatttaagt	1380
aagatttagg agatttat ttttttttta gtttaggtta aattattttg gaataaataa	1440
aattttgtttt attttaatta ttaatattgtt aaaaagtatt aagtagttt tatttaagta	1500
atatttaat attttgatata taaaaaaatg tttttgtat atgaaattta ttatataaaa	1560
taaggttagat aggttgtat attggtagt tatgataata ttggagttt gtaattggaa	1620
gattttttaa aaggaaattta ggggatttt ttttagattt gtttagtttta taatggtag	1680
aattttatgtt aaattttggta taatttaatg atatttgagg aggaaggggg gttttttagg	1740
tagggattat ttataaaagt ttttgggtga aaaaaatttga gttttgtgtg ttaatttttag	1800
gtttggttat gatthaattt ttgtttttt attttaattt attttataat ttgttaatga	1860
atatttataa ttgtttttaa tataatttt agtgatattt ttaggattaa taaaaaaagg	1920
tatgttaaaa ataaaagtat tatgtaaatg taagttatatt tatgaaaata tataatgtat	1980
ataatttttt tagtaagata tagggttt ttagtattttt gattttttt ttgattaatg	2040
ggttaaggggtt gaagaagttaa tggtagttaa ggagatattt taaaaataaa ggaaaaattt	2100
ataggagtga ttattttttt gttttatattt tttttataaa gttaggtttga aaatatttag	2160
tatattataa attatatgtt agaggttaggg atttgataga atttgaataa tgtgaatttt	2220
aaaattttttt attttaataa aaattaatag gtaattttat attaaaataaa taaaataaaa	2280
taagagaaaa ggttagtaata gaaaaaaa tgggtatgtta taagttattt gagatataga	2340
agttataattt ttatataattt ataaaaagag ttggatgggt aagatgagta gagattgtta	2400
aaagtattttt ttattatagg aataaaaaaa ttaattttt agatgaataa tttaattttt	2460
aattttttttt	2470

```

<210> SEQ ID NO 15
<211> LENGTH: 2229
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

<400> SEQUENCE: 15	
ttttttttt gttttgggtt gttttgggtt gggtttaggtg gagaagttgtt tttttgttaa	60
ggtgtatagaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	120
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	180
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	240

-continued

<210> SEQ ID NO 16
<211> LENGTH: 2229
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

-continued

<400> SEQUENCE: 16

tttttttttt	ttttttgtt	tttgatttgg	tggtgatagg	tttttgtga	aagttagattg	60
ttattttttt	gtttatattat	ttgtaaaagt	agtgtttta	gtgttaaggt	ttggggaggt	120
gtgggtttagg	ttggaggttt	tgggggttgc	tttaggggtt	tagtttgc	ttttgaatt	180
agattttagt	gttaaatattt	agagggtatt	tatTTTTT	gtattgat	aatttatgt	240
ttatgaaatg	tgtaaatttt	gttgggggttgc	tatTTGaaatt	tagttagaga	attgggggttgc	300
gttttgtata	taagagatga	tttaaagaag	gtagaaatga	aaatttgata	gaagtagtaa	360
tagttgttgc	gggtgattgt	gggggtgatttgc	taggaaaatttgc	tgagttttttgc	attgaagtttgc	420
gttttaggaag	gggtgattgt	tgttaatttgc	tggagggagg	gatgggtgaag	gaagatgtttgc	480
ttaatTTTTT	atTTTgttaa	tttttatagt	ttgttttttgc	agtttgcgtt	aaatataaaaa	540
gttgttaaat	agaaaaatag	agggtttgg	ttaaaatttgc	atattttaa	ttaatttgatttgc	600
tgtatgattt	tgtataatttgc	tttttttatttgc	ataattttata	tatagtgaag	tgttttagtttgc	660
tgaggaggaa	agtttggaaaga	attgttatgt	atgtgtat	tgttagtttag	gtgtattttgc	720
ttgataaaga	aatagattat	ttatatgtttgc	ttttaaatgt	tgttaaatgt	taaagtaaaaa	780
atatTTTTT	taagttgtttgc	taagaaaatttgc	ttagaagata	ttttggagta	ttggggatttgc	840
tatTTTgttt	ttgatagtag	tgttttttgc	atagggtgat	atTTTgttta	gggtattttgc	900
gtgggtttgc	ttttaatgt	aagtttaag	tgggtgtgttgc	aggaattttgc	ttgtgattgt	960
tgggttaagt	tggagttatttgc	aagttttgc	ttgttatgtgt	tgtgtatgttgc	gtatgtatgt	1020
ggaaaataga	ttaaaatgttgc	ttatagaaaa	ttttggtaa	gttttggagg	atTTTgtttgc	1080
ttaagattag	ttgggtgtat	tttttttgg	atgttttttgc	tttttgggttgc	tagtgtttttgc	1140
ttgtttttat	ttgtgtgtat	ttttgc	gtggtaaatttgc	gaaataagaa	atggaaatatttgc	1200
attgggttttgc	ttgtgtgtat	gggatgagag	gttgcgtat	tttgcgtgttgc	ttgtttgggt	1260
tttgggttttgc	tttgcgtat	tttgcgttgc	ggtgcgttgc	gttgcgtat	gagggggata	1320
gttgcgttgc	gtttttgtgg	tttagaggttgc	ttgggtat	ttatgggggg	gggtgttttgc	1380
gttttgcgttgc	gtttttgttgc	ttgagggtat	gtgagggtgg	tgggtagggg	agtgtatgttgc	1440
ggagagaaga	gagaatgtttgc	aatttagagag	aattttttgt	ttttggatgttgc	tttagtgcgttgc	1500
tagtttttgc	aattttgttgc	taggaagttgc	aaggatttgc	gtttaggttgc	ttgtttgggg	1560
ttgggtgtat	agagtgttgc	tttgcgtat	gtgtgttgc	gttaggttgc	atTTTgtttgc	1620
gtgtgtgttgc	agtgttaggttgc	tttttgcgttgc	ttgtatagg	ttttatgt	gggtatgttgc	1680
atttggatttgc	gaaattaata	ttgtgaggat	ttgggtatgt	gtatgtat	tttggggaa	1740
tttgcgtat	agttgattgttgc	tttgcgtgttgc	tattagat	gtgagaagta	gtttttttaa	1800
atagggttttgc	ttgttggaa	gaggaggtat	ttttttat	ttgtgagtt	tttttagttgc	1860
taattgggttgc	ttttttat	attagttat	gtgggttttgc	gagggtatgt	ttggatatttgc	1920
ttttttgttgc	agatgtataa	attggggata	tttttatttgc	tatTTAAGAT	gttttttttgc	1980
tatTTAGTAG	aggggtgtgg	agtaaatttgc	ggataataat	ttgtgtgttgc	tttgcgtat	2040
gtttttttat	aaaggatgtat	aaaaatttgc	tgtgtggaa	gaagtttttgc	ttgtgtttat	2100
gaaagggtat	ttgtttttat	ttgttatttgc	ttttgggttgc	ggtttttatttgc	tttagtgcgttgc	2160
tttgcgtat	taataaagag	ttgtttttat	atttgcgtat	aatttgcgtat	agttgtatgttgc	2220

-continued

gaggaaaga	2229
<210> SEQ ID NO 17	
<211> LENGTH: 7833	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)	
<400> SEQUENCE: 17	
gttttggtg agatatgtgt tttataagtt ttaatggaga aaaatgttaag tattttat	60
tttgaattt ggttatttga gtaatgagaa aatagtattt ttttttagga tagtgggttt	120
taattatggt tatgtgtttt tttaggaaaaa ttttaaaaat atatataatat taatgtttt	180
gtgttatttt tagggatttt aagtttttga atatgaattt tgtatttagta ttttttaatt	240
atttaggta ttgtgatgtg aaattatgt tgagtttat tgtttaaga tgaaataaat	300
tttttttagt attgaaatta taaatttaaa ttattnaaat taattaaggg tatggaaatt	360
aataaggat agggaaagttt ttatattata aaatttattt tttaaattat agtttattgt	420
ttatatgtta ttgttattt tagaaaaggg tgaaaaataa gttaattttaa ttattnnag	480
tttgaaaaat tatttagaaa tgaagatgtat gattttgaaa tattgttaat attatttgat	540
ttataaataa tgtttaata tattttattt atattgtatg atattttttt atatgaat	600
tatataattaa aattaaggta ataatgtatt tagaatattt tattttatatt tatgttttt	660
aaggtaggtt gaaattaaga tatgagttat taagtatgag atgttaagggt gtggggtag	720
aaattatattt gtattnattt attaataattt aatataattt ttaatattt atatattttaa	780
tttaattttg tatattttta attattnntt attatgtgtta taaatataag tataatattt	840
tttatgtatt tattttattt tattttattt tattttatattt tataagggtt ttttttaat	900
tttattttat taaattttat attttttttt taattnntt tagttttttt aataaggttt ggaggttagt	960
attgttattt tttatattttt ataaatgagg aaattgttta tagttataaa gttattgtgt	1020
tagatattt agaagtttaa tatatattttt gtgaatataat gtataaaaat agagagat	1080
atatgtataa tagtttattt ttatatttgag taaaagtttt taattgttt tagaaatttt	1140
tttgtgaaaa ttgagttaaa attgaggat tttttttttt gttatataagg tataagggtt	1200
attttatttt ttaataaagg atgaatattt aatgtggat ttttaagggtt aatttttagat	1260
tttttgaattt ttgtatgtg ggattttggaa ttgtttttt gttttaaagt ttttttaagg	1320
attttatgtt ttaatttaggt ttagaaatttta ttggattttt ttgttgaagt ttgagaattt	1380
aagtttgggt ttatgtgg tttatagaa agggttaatgtt aagtattatg gatagaattt	1440
atatgtttttt agtttagttt ttttttttaga agttaatagg tagtaatata gtagaaattt	1500
gtgattttatg ttttgggttt tgaagtttagg tagaatttttta tagagttttt gtagtggat	1560
tgatgagattt ttttttttttgg ggtttaggtt ttgtatgtttt taaagttata tttttttat	1620
ataaaatgag ataaatatttt ttgttttataa ggggtgtttt aaagattaaa taaaataat	1680
atgtttttattt ttatatggta taatgtttga tatttaagaa gtaaggata tattttattt	1740
tattttagt aatttagaaag tatgaaatttta tgaaggat aagagttttt atggtagt	1800
tattttattt ttttaggtttt atttattttat tttaaattt ttttgggttga gaataatttt	1860
taagttttttt attaagttt gtagtaattt tattttttat aatgtatgtt tttttatgt	1920

-continued

aaaaaaaaat gttttaagt ttttggaga aaatatattt gtattatttt tattaaaaaa	1980
ttaataatt ggattttgtt ttttgtatt aatttttagag tgtatatgtt ataaataaag	2040
tgttttagtt taagaagatt gaaagtaaat atggatagt attttaaat aagaattttg	2100
taaatatatg gtatgattgt gttatattat tagtaattat atgatatgta atgtaaagta	2160
tagttatag atttaaattt aatttaata agtaaattga tttgttttgg tttggggaaaa	2220
gttaaagtat taatttaattt gtaatgttag tttgtttat tttttggta tttagtgata	2280
agtttaaata atgtatataat ttttattttat atatttagta atataatttt ttgtttaaatg	2340
agtgtatgtt tttgtttattt tggtgggtt tgtagttt agaatttgg tttgtgggt	2400
attataataat taagtataga gtaagtgtaa taaaattgta gtattttat tgaaaagggtt	2460
ttgttttaaa ttgtttataa atttaaaggaa tttttgtga agtaattgta ttgtttaaatt	2520
agttataatt agtaatttaat tttttggag tttaattta tttttggtaa aatgttttag	2580
gaagagtata tattattaga aagtatgtta aaaatttt tagtagaaaa tttaaaaataa	2640
gtttttttt gtttaagaggt ttttaaaat ttttattata tagttaaattt ttgaaatttt	2700
agtaggtttt gttttattat tataattttt gtataaaatat ttttaaggat ttgtttttta	2760
gttttaagta tgattttattt ttataagttt gatttagttt tatattagtt ttgtttatgga	2820
aaatgatatg tttttatttt ttgttgtaga gttgttaat tttgattttt attttatgtt	2880
tttttttgt tgaaagttt tagtgaaga aatttttaat tttttgtttt gtaatattag	2940
ttggtagttt tatttaatgg gtattttgtt ttttaaaga aattttgtt tttgttttaga	3000
agttgatttt ttgtatgttt taatgtttgg tttaattgtat ttgttttaat ggagttttg	3060
ttggtgagga gtgagatgtt attgattaga atgtgggat ttgttggtta attgttagga	3120
gtgagagata ttgagattta gaaatttttgg gaggtgggag gggagagggaa tagttttgga	3180
tggaggttga gatgttaatgaa aagggttgg attttatata gaaaaaaa aagattttg	3240
ttgaggtatt gaggtgttgt atgattataat ttttaaagg agaagttaaa aagtaaggaa	3300
gtggggaggag gttggaggtt aaagtattta aaaggattat ttgggttataa ttgtttttt	3360
ttgttggttt tgtaaaggat agatgtttt gtttttaaag tatgtatgtt ttttttttaa	3420
gtgattggga atggatattta attgtttgtt aaatgttattt aaatgtttt tttaatttag	3480
gggatataga aagaggggta taaaaggaga atttaatag aaaaaggag gatttgagg	3540
ttttgaaag tggggggaga agaaggagga gggataatag agaggaatag agaaggagag	3600
tggagagaag ataataaaaa ataaaaatag gaattattga ataatttat attaaaaaga	3660
aagttttttt ttatgggtta tttaaaatat tgagattgtt atgtgtt tggttatgga	3720
agaaagatgt tttttttat tttgtttttt gaaagttttt gggtttgtta ttgtgttatta	3780
aaattttattt aggttaaaga gtgtgtttaa ttgtttgaag aatgttagtag atggaaagggt	3840
ggttttgtta tggtgtttgt tttttttgtt ggagagaatg aaagaaatgt gtagagttag	3900
agattttgtt tgagtttagat tttttttgtt tgtttttagt tattgttatt ttgtttaaaga	3960
tttgagtaag gaatgttaggg ttattgttttgg ggttaataaaa tggagtttgg tttttttt	4020
ttggatgttg ttgtttgggtt gatgttttttggtaatttatt tggtgttat gtagaggagt	4080
ttttttttt ttttagattt atttgttttg attaatttga ttttttaat atatttgattt	4140
gtatttttta ggtggatata ttaataggtt atgggttgga gaggagttgg tgatgaggag	4200

-continued

agggatttaa atttgtgaat gtttgggttg gggtggagtt gtgggggggtt tgggaggaga	4260
gaggggagaa gagagaagga aggagagtgt ttgttggat gggtgagttg ttttggtag	4320
tagtttggg gttgtatgtt tttgtggag atgttggtgt tggttttagg ttggtaagag	4380
tggtttaat attattgttt ttatattttt tttttgtaaa ttttttagaga aatgtttttg	4440
gttttttgt tttgtatgtt ttatgttata ttttttata gtttaggtgg tttgttttg	4500
tatgttggag tttgtgggtgt tagtaggtat ttttttttg tttgttttg ttttttttg	4560
ttttgttggttt gttgtttttt tgatattttt gtttttataa ttttttagttt ggagagatgt	4620
tattnatgtt tttgtttgtat ttgtgggttg gggttatgtg tggagggggg gtgttagttt	4680
ggattnatgtt tttgttaggg ggtgtttgg agtggagagt gaggtaatg gtatatgagt	4740
gtgtgggttag tttatatttga agtggatgtt tttatatttga gttatgtttt gtttagttt	4800
attnatggtaa gtgtttgggtg agtggatgtt tttgtggttt ttgtgggtgt ttttttttg	4860
tattnatgtt tttatatttggttt gattnatttt ttttgggtt ttttgggtt ttttgggtt	4920
agattnatgtt tatgtgattt tgggtgggtt gaggattgt tgtnnnnnn	4980
atgttagagg gatgtgggtg gttaggtttt agaggtgggtt gttgggtgtt ggggttttt	5040
gattnattnn tttatatttggttt ttttgggtt tttatatttggttt gtttttggat ttttgggtt	5100
ttttgtttttt tgggtttttt gattnattnn tttatatttggttt tagtttttggat tgtnnnnn	5160
gtattnattnn agtggatgtt tttatatttggttt tgggtttttt gttgggtgtt gatgttttt	5220
gaaattnattnn agattnatgtt tgggtttttt gttgtttttt ttttgggtt ttttgggtt	5280
tgtgtttttt tttatatttggttt tttatatttggttt gatgtttttt gatgtttttt	5340
tattnatgtt tttatatttggaa gttgtttttt gttgttaggt tttatatttggat tttatatttggat	5400
tgtgtttttt tttatatttggaa gttgtttttt gttgttaggt tttatatttggat tttatatttggat	5460
tttatatttggat tttatatttggat tttatatttggat tttatatttggat tttatatttggat	5520
attnattnn gttgtttttt tttatatttggat tttatatttggat tttatatttggat tttatatttggat	5580
tttatatttggat tttatatttggat tttatatttggat tttatatttggat tttatatttggat	5640
tttatatttggat tttatatttggat tttatatttggat tttatatttggat tttatatttggat	5700
tttatatttggat tttatatttggat tttatatttggat tttatatttggat tttatatttggat	5760
gggttagaaag tttatatttggat tttatatttggat tttatatttggat tttatatttggat	5820
gaggattnattnn agattnattnn tttatatttggat tttatatttggat tttatatttggat	5880
taattnattnn gttgtttttt tttatatttggat tttatatttggat tttatatttggat	5940
aggatgagga agattnattnn tttatatttggat tttatatttggat tttatatttggat	6000
tgtgtttttt tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6060
ttgtgtttttt tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6120
ttttttttt tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6180
ttattnattnn tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6240
ttttgtttttt tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6300
ttttttttt tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6360
ggtggattnn tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6420
ggttaggtat tttatatttggat tttatatttggat tttatatttggat tttatatttggat	6480

-continued

ggttttgttt tggttaagtt tttaattaa atagggttag aggtgggag ttgttgttatt	6540
ttagtgtt atagtattt gtttgatagt ttgttagata gggtgtatgt aatttttat	6600
ttttgtgaa tataatttt ttgttagttaa atttggttt gaataaagtg tttttaaag	6660
atgtatataa gttgaagtgt atgtatattt agagaggagg gaatgattaa ttgtatattt	6720
gggtgaaagt ttgtatagtt tttagttattt attgtatgtt atgttaaaag gaaaattattt	6780
atgtattattt ttaattttt ttttataaag ataagttgag atatgttaatt ttatttagatt	6840
tgggttaata gattgtttt tttttggta gtttttaat ttgtatattt aataaaattt	6900
aatatgtttt tataattttt tgattttatgt gtatatgtgt gttgttttg aaagaataaag	6960
ttttatttt ttattgttta attatttttt agatgttttta ttatgttaat aattatgagt	7020
ttgtaaaaat aatttttggaa aatgttgatg gttttgtatg ttaatataaga ttggtttgg	7080
ttatttttat gttttgttattt gtttttagaa ataattaatt taaatgtgaa gttgatattt	7140
gtaattaaga aattatataat ttatttagata ttttaaaggg gattgtataa attaaagaga	7200
ataaatttggt tttgttagata ggttggtaag aatttggat tttgtttta tttttgttaa	7260
tttagaggtg attatttttt atttgatgtt aatagattat tatagaaaaat attgttttg	7320
tttattttta ttatttgaggt tttgtttttt ttttggggat atatattttta aataaggggt	7380
tggttttagttt gttgaagttaa aagaataattt aaagatgggg aaatggtaaa agggttattt	7440
gagattttaa ttatgtttttt ttttaaatgtt ggagttttgtt ggttataat attgttttattt	7500
taatgtataaa aaaataaaaaaa taaaaaaaaa ataggaagta aatgttaatg ttttattttat	7560
tattgttagt attaatgtaa gttttaaaaaa atagtattat tagaaaagga tattaaagga	7620
gaattgatta gaaaagaattt gtggaaaatgtt gaaatgataa ttgattttt aatttagattt	7680
tgaggttattt agtagatgtt gttttgtatg tataatgttata gttgttgat tttaaaatttta	7740
ggataagttat tttaaagttt taaaatgtt gttttttttt tttttttttt tttttttttt gtaagatgtt	7800
ttaatgattt ggttgggtt tttgaattttt agg	7833

```

<210> SEQ ID NO 18
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

<400> SEQUENCE: 18	
ttttaaattt aaagagaata tttagttat taaaatattt tatagatttt taataaaaaaa	60
aagtattattt ttgaagttttt aaaatattttt tttaaaattt taaatttaat aattatgtt	120
gtattgtttaa gttttttttt ttatgtttaa tttttttttta gtatttttt ttgtatgtt	180
tatatttttaa agttgtgtttt aatattgata gtggtaatg aaagtttaat atttggggat	240
tgtttttttttt ttatgtttat tttttgtttaa ttaggtggat aatattttatg attataaaaat	300
tttatattttt ggaaaagagt tagtgtatgat ttttgaatat ttttttattt tttttttttt	360
ttaatgtttt tttttgtttt aatgattgaa ataattttttt atttggaaatg tatttagata	420
aagaggaaat aaagttttaa taataaagat aaataggtat agtggggat gtgtatgtttt	480
gtttggttta aatgaagattt gattttttt aagttaatag ggggtggaaat ggggtgttaa	540
	600

-continued

-continued

ttggggttat	gtgggaagag	gtagttgggt	tttgcatttgtt	ggagtaggat	gtaggttttg	2940
ggagggaggg	gttgatgagt	aggtgttaagg	atgttaaggag	gagggtggttt	tggaaagttat	3000
agatgggtt	gtttgttagg	tgttggttt	agtggggta	ggtggggat	ggtttaaatg	3060
agaagtttg	gttttagggt	gggttatttt	tatattttata	tatttttgtt	tttatttttt	3120
gttttaggat	gtttttattt	gaagggtgggg	tttggattag	tgtttttttt	ttgtgtgtga	3180
ttttgggtt	tgagtgtggg	ttgtgggtt	gtgggtttt	tttgagttgg	agatgggtgg	3240
ggtggaggtg	ttagaggagt	agtagtagta	gggttagagag	gggtgagttt	gtgtgggaga	3300
gggtgtttt	ttgggtattt	gtgttttttt	gtgttttttt	gtgttttttt	ggttgttaggg	3360
ggatgttagt	tggaaatgtt	gtgggtggaga	ggtttagggat	gttttttttt	ggatttata	3420
gaaagagggt	gagaggttat	ggtgtttagaa	ttgtttttgt	tgattttggaa	gtatagtag	3480
tatTTTTT	aagagtgtgt	aatttttaagg	ttgtttgtt	aggtgtttta	gttattttgg	3540
taggtgtttt	tttttttttt	tttttttttt	tttttttttt	tttaggtttt	tgtagttttt	3600
atTTAGTTA	agtgtttgt	ggtttgaatt	ttttttttta	ttatTTTT	tttttttagtt	3660
tgttagttt	tagtgtgttt	atttgggggg	tgtgggttta	tgtgttttgg	aggttagatt	3720
ggttgggata	agtgggttga	gagaaagaga	aaggTTTTT	tgtatatgtt	gtgggtgggt	3780
tgtggggagt	attgggtggg	tagtgggtt	tgggaagggg	agagtgggtt	ttatTTTTG	3840
gtttaggtag	tgatTTTTG	tttttttttt	gggtttttgt	tggatgggtt	gtgatttggg	3900
gtgtgagag	aaggTTTAAT	ttggtaggag	ttttgggtt	tgtgtttttt	ttttatTTTT	3960
tttagtggga	agggttaatg	gtatagtggg	atttggTTTT	tgttgggtt	atTTTTAGG	4020
tagttagata	tatTTTTAG	ttaatggaa	tttttagttt	tagtaatggg	attaagagtt	4080
tttggggata	agggtggaga	ggaatatTTT	ttttttatga	ttggggat	tatttttagtt	4140
tttagtggttt	ggatgtttta	tagggaaagag	tttttttttt	gggtgtgtat	tatTTTTG	4200
tttttggTTT	tgtttttgtt	tatTTTTT	ttgttttttt	ttttatTTTT	ttttgttat	4260
tttttttttt	ttgtttttaa	aagttttgg	attttttttt	ttttatTTTA	tttttttttt	4320
atTTTTTTT	ttgtgttttt	ttttttgtt	tttttgaatt	taggagagta	tttgataata	4380
ttaataggt	aatttagtgtt	tatTTTTAAT	tatTTAAAG	aggtattttat	atatTTTGAA	4440
aatgggatta	tttatttttt	gtagatatta	gtaaaaaat	aaattgtatt	ttagtaattt	4500
ttaatagtat	ttaatTTTT	aatttttttt	tatTTTTG	tttttaatt	tttttttga	4560
gagatgtgat	tgtgtgttat	tttagtgttt	taatggaaatt	tttttttttt	ttttgtgtga	4620
aattttatTTT	tttattttat	attttgttt	ttgtttgaga	ttgttttttt	ttttttttat	4680
ttttaaagat	ttttgaattt	tagtgtttt	tatTTTTG	aattaatgt	tagattttag	4740
tatTTTTAGT	gggtgttattt	tgtttttat	tgtatgaagat	tttattttaa	tagattaatt	4800
agatttagatg	ttggaggtat	tagaaaattt	gttttttagt	agagttagta	aatttttaa	4860
gaaaatagaa	tatTTTTAG	atagagtgtt	taatggaaat	tgtaaaataa	ggaatttagaa	4920
atTTTTTTG	ttataggTTT	ttagtagaga	aggtaatata	aatatagatt	aagattaat	4980
aattttatAG	tagagaatga	gaatatgtt	ttttttatag	taaggTTGG	gtggtaatta	5040
attaggttta	tgaaaataag	ttatgtttga	aattaaaggt	aaagtTTTA	aaagtgttta	5100
tgttagtaatt	atgataatga	aataggattt	gttaggattt	taggtttgg	ttatgttaatg	5160

-continued

agaattttag agaattttt agtagaggaa aattgtttt gaatttttg ttaagtaat	5220
ttttggata tttttaata atatatgttt ttttaagat gtttgttaa aagtaagtt	5280
aaatttaaa ggagtttaatt attgggtgta attggtaat aaatgtggtt gttttatag	5340
agggttttaa aattattaaa tagttgaag taaagtttt ttaatggaa tggtaatt	5400
ttgttgtatt tattttgtat tttagtgttat agtgttatta agaaataaat tttgaaattt	5460
gttaagtatta ttaagtggta gaagaatattt atttattgag tagagaatgg tattattgaa	5520
tatgtaaata aaaatataata tattattnn atttgttatt aggtattaaa gaagtagata	5580
agattgtatt agtaatttggaa tttagtgtttt aattttttt tagtaaggta aaatttagtt	5640
atttatttata attaaattta agtttatgaa ttgtattttg tattgtgtat tatatgattt	5700
tttagtaatata gatataattta tattatgtat ttgtaaaattt tttattnaa aatattatata	5760
tatattttt ttaattttt ttgagttaga atattttatt tttggatata atattttaga	5820
attgtatgtttag aggagtagag tttagtggat agattttttta gttagaaatag ttttagatata	5880
tttttttag aaaattnaa aatattttt tttttatgg aaagaatattt attataaagt	5940
gttagattat ttatagttta agtaggggggt ttggggatg tttttataa agaataatgtt	6000
aagataaaata aatgaatttggaa gggaaataag atatattgtt aatttagaaattt ttttttttt	6060
ttatgatttt atattttttgg attgttttaaa taaaggtaag atgtttttt tgtttttag	6120
gtgttagtta ttgtgttattt taggatagaa tatgttattt ttattnaaattt tttaaaatata	6180
ttttatgaga taaagaatata tattttattt tatataaaaag gaatatgtt ttgaaatgtt	6240
taggttaattt gttttaaagaa ataaattnn tttagtgatata tggtggatt ttgtgaaattt	6300
ttgtttgatt tttagtata agatataatgtt tattatgtt tggtgttattt tggtttgtt	6360
gtttttgaga gggaaatata atggaaatgtt tattatgtt gttttagata ttttattgtt	6420
ttttttgtt gagttgtatg aaggtttaaa tttaattttt taaattnn ggaaatgtt	6480
tagtgatttt tgaattttggat tgattatgtt aatttttggaa gaaattttggaa aataatgtt	6540
aaattnaaat ttttattttt agggatttttggaa aattnn gttgggtttt gggatttata	6600
tttaatattt tttttgtt ggagaatgtt ggttattttt atattttatata gataaaatgtt	6660
aggatattttt gttttttgtt tagttttatgtt agagaggtttt tgtagatagg ttaaaatgtt	6720
ttattnatgtt taaagatgtt ttgtgttata tggtttttt tttgttttta tgtagatgtt	6780
tattaaatata gtattatgtt ttatgtatgtt aatttttggaa ttgttagataa	6840
ttttttattt tgtaaaatgtt gagtaataatgtt aatattttgtt ttttgggtttt gtttaaaagaa	6900
ttaaaatataa aatgtatgtt ttaatgtttagt tggattttggg gggattttttt atataaataaa	6960
gtgaatggag gtatgtatataa ataaatataatgtt aatgtgtt gttttatataatgtt	7020
aattaaaaat agttaaagat gtataaattta aagttaaatgtt tatgtgtatata tttttatataatgtt	7080
gttgattttt gataatgtttagt tttttatgtt aatattttgtt tttttatataatgtt	7140
ttaataatgtt atattnaaat tttttatgtt tttttatataatgtt aatattttatataatgtt	7200
ttaaaatgtt tatttttttta gttttatgtt aatattttgtt tttttatataatgtt	7260
tgtgttagtta atgttattttgtt aatattttgtt tttttatataatgtt aatattttatataatgtt	7320
gttgttattt ttattnnn gttttatgtt aatattttgtt tttttatataatgtt	7380
ttattnnn tttttatataatgtt aatattttgtt tttttatataatgtt aatattttatataatgtt	7440

-continued

ttttataatg taaaagttt tttatgtttt attgattttt atgaaaaaa ttaatttgg	7500
tagtttaagt ttgtgatttt agtgttgagg aaagtttatt ttatTTTaaa gtagtggggt	7560
ttagttatga ttttatatta taatttattt gataattaaa gaatattgt gtagagttt	7620
tatTTaaaga tttggaaattt tttagaagtga tatagaagta ttggatata tataTTTTA	7680
aagTTTTTTT ggagaaatat atagttatga ttgagaatta ttgtttggg gaaagtgtt	7740
atTTTTTtat tatTTaaata gttaaagtttt aggaggtaaa atatTTATat tttttttat	7800
taaaatttgt aaaatataata ttttattaaa gat	7833

<210> SEQ ID NO 19	
<211> LENGTH: 5666	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)	
<400> SEQUENCE: 19	
aaaatttagaa tttttatTTTT tttgtgtttt ttatTTTT tagtgttttt taatTTTT	60
ttgttaagtga ggggtgggaa ggggttttat aatTTTTta gggagtaagt tttttttgg	120
ttttttttt tttttttttt tttttttttt tgagattaag tttgtttttt gtttttttag	180
ttggagtgtta atgggtgtat tttggTTTT tgtaatTTTT gttttttttt gggTTTaaGT	240
gatTTTTta tattagTTTT tgtagtagttt ggattatagg tatgtgttat taagTTTTgt	300
taatTTTgtta ttttttagta gagatagggtt tttgtttatgt tggttaggtt tgTTTTgaat	360
ttttggTTTT aggtgatttt tttgtttggg ttttttagaa tgTTGGGATT atagatgtga	420
gttattgtat ttggatTTTT tttttatgtta atagtgtataa ttttattaa agtTTTTTT	480
tttttttttT agttggagtt ttatTTTTt attttaggtt gagggtgggtg gtgtgatttt	540
gtttttatgt aatTTTTgtt ttttggTTTT aagtgtattt tttgttttag ttttttaggt	600
agttggaaattt atatATgtgt gttattatgg ttagttaatt tttgtatTTTT tagtagagat	660
ggggTgttat tttttggTTT aagtggTTT tgaatttttT attttaggtt atttgtttgt	720
tttggTTTT taaagtgttgg ggattatagg tttgtgttat tttgtttgtt tttaaagtat	780
tttttttttta tttttttaaa taagattgtta agtttagTTTT taaagtggat aatTTaaagag	840
ttaataggta ttagtttagg atgtgtggta ttgtttttaa ggtttatATg tattaatata	900
ttatTTAAat ttataataat ttttataaaag tagggggatTT ttagtattttt tttttttttt	960
ataattatga aaaatgtaaat gtatTTTTat taggaaagag aaatgtgaga agtgtgaagg	1020
agataggata gtatTTGAAG ttggTTTTTt gattattgtg taatTTTTgtt tttagaaat	1080
ttagtattttt ttttggTTTA ggaattatgtta ttttgagaat ggagTTTTt ttttaatgt	1140
tttttttttT atTTTTTTat ttgttttagt gttagattttt tttttgtttt gttttaaataa	1200
atTTTTTTTt ttttagagTTT gttttatataat taggtatgt atatgttga gaaatTTTTG	1260
tttttagatag ttgttttataat ttgtgggggg gaaaggggagg ggaaggaggag agtagTTGA	1320
ttttttaaa ggaattttttt gaatttaggtt ttttgatttta gtaatttttT gtttttgaa	1380
aattaagggt tgagggggta gggggatattt ttttagttgt ataggtgattt ttgttttttG	1440
gtggggTTTT tataattagg aaagaatagt ttgtttttt tttatgttta aaagaagaag	1500
ttatTTTTTTt tttatgtatata taaatTTTTt gatttaattt ggttagttagg aaggTTTgtat	1560

-continued

tgtggaggaa ggaaatgggg tgggggtgga	tttttttta atagagtcaa	tgtattaaa	1620			
tatgtttttg ttggtaggtg	ggggagtg	tttggagta	gggaggttgg	agggttgtgt	1680	
gggggtttagg	tggggaggag	tttagttttt	ttttttgtt	aatgttggtt	ttggtgaggg	1740
ttgtttttgg	ttgggttttt	tggggagat	ttaatttggg	gtgattttag	gggtgttata	1800
tttggtaagt	ttttggagtt	aatagtattt	tttttgagta	tttggttatg	gtgtttttt	1860
gtttggaaag	atattgtgtt	tttttttagag	gatttgaggg	atagggttgg	aggggggtttt	1920
tttggtagta	ttggaggaag	aaagaggagg	ggttgggttgg	ttattagagg	gtgggggtgga	1980
ttgtgtgtgt	ttgggtgttgg	tggagaggggg	gagagttagt	agtgggttgtt	ggggagtagt	2040
atggagttgg	ttggggggag	tagtatggag	tttttggttgg	attgggttgtt	tatgggtgt	2100
gtttgggggtt	gggttagagga	gggtgtgggtt	ttgttggagg	tgggggtgtt	gttaatgt	2160
ttgaatagtt	atgggtggag	gttgatttag	gtgggttagag	ggttgttagt	gggagtaggg	2220
gatgggtgggt	gattttggag	gatgaagttt	gtaggggaat	tggaaattagg	tagtgtttt	2280
attttttggaa	aaaagggggag	gttttttggg	gagtttttag	aaggggtttt	taattataga	2340
tttttttttg	gtgatgtttt	gggggtttgg	gaagttaaagg	aagaggaatg	aggagttagt	2400
tgtgtataga	tttttgaat	gttgagaaga	tttgaagggg	ggaatatattt	tgtatttagat	2460
ggaagtatgt	tttttattag	atataaaattt	tatgaatgtt	tggataaaaa	agggagtttt	2520
aaagaaatgt	aagatgtgtt	gggatttattt	agtttttaat	ttatagatat	ttggatggag	2580
tttatttttt	ttatttaggag	ggatttattag	tggaaatttg	ttgtgtatgt	tggaaataat	2640
attgaatata	aattttgatt	gaaatttattt	agaagtggtt	gggtgtgggt	ttttatgttt	2700
tgtatattttt	ttatTTTGGG	agattaaggt	ggggggaaattt	atTTGAGGT	gggagtttga	2760
gattagtttg	gttaataggt	gaaattttgt	tttttattaaa	aatataaaaa	gtatTTGGG	2820
gtgggtgttag	gtgtttgtaa	ttttagttat	ttggggaggtt	gaggttaggag	aattgtttga	2880
atTTGGGAGG	ttgaggttgt	agtgaatagt	gagatggagt	tattttattt	tagttgggt	2940
gatagagtga	gattttgttg	aaagaaagaa	agagagaaag	agagagagaa	aaattattta	3000
gaagtaatta	tatattgtgt	ttatTTTAA	ttgagtaggg	taaataaata	tatgtttgtt	3060
gttaggaattt	aggaaataat	gagtttatattt	tatgtgatta	ttttagaggt	aatatgtagt	3120
tattatTTTG	ggaatatttg	ttaatatttt	tgTTTTTTA	ttatTTTTAG	tttatttgat	3180
atagttttat	tgtgataaga	gttttaattt	ttttatTTTT	gaatagaggt	gttttttttt	3240
ttttttatTTT	tgTTTTGTGA	gggagtttag	ggaggattta	aaagttaatta	atataatgggt	3300
aatttatgtat	ttttaaaattt	ttgttaatag	tttgaatttg	ggagtttgtt	ttttagttt	3360
tataatattt	tagaagagat	tttatttgg	taaaaataaa	aaggaaaaag	aaaagtggat	3420
agttttgata	atTTTAATG	gagaaggggag	aagaatatgt	agaaaaagggg	aaatgtgtt	3480
gttttagaat	ttaattata	ttgggtgttta	atataggaaat	atttatttat	ataatatttt	3540
aaagtattaa	atttataattt	gtatattattt	aaatggatat	attttaat	gggtttaagt	3600
atTTTatata	tttaattta	attgatttat	tttttttttg	ttttggattt	ttattatgtat	3660
ttaaatattt	atataatgggt	tattttttag	ttttttataa	ttatgaaata	taagaaaaat	3720
tttaagggtt	agttttatga	ttaagatgaa	ggattttattt	gaatataaa	aataataat	3780
atattgtat	atTTTgtttt	ttttttgtta	gttgaatattt	ggttggtttta	tatTTTTTTT	3840

-continued

ttgtttttt gaaaattgag ttagtttat ttttttagga taggattaa taattataat	3900
ataatttagt ataattttt gathtagta aattatgtaa tttgtgtta gtatgaaatg	3960
tattnaaaaa taagtaattt tttttaata ttattatTT taaattaata taataaataa	4020
tagttatTTT aaaataaattt gtttattttt attatgtagt atttaaattt taaggttgtt	4080
atgattgttag atagtatttt aaaattttt tttggaaatg gttttgttt taagatgatt	4140
taggaattaa agaggtgatt atttttgtt taatgaattt ttaaattata aattttggaa	4200
gtgttttagt tttttattgt tggtgtata aattattata aatgtgttag ttAAAataaa	4260
tataaaatta ttatTTata gtttttagaga ttagaagtta aaaatgggtt tataaggTTT	4320
tatTTTTT ggaaatttta agggtaatt tgTTTTTtg ttttttagt tgTTTgtga	4380
ttattaaattt ttTggTTta tggttttgT atTTTTTtg tggtttgtg ttTTTTTTT	4440
gtatTTTTT ttgattgtg atTTTTtaat aaaaatattt ggggttatgt tggtttatt	4500
ttgaaaattt tggataattt ttTTtaagat tattaattaa attatTTtg taaagtTTT	4560
tttggTatAt aagttaatgt attaaaagt tttgaggatt aggatataga tattgggggt	4620
gggggggtat tatttagTTT attatagaa ggaattttag ggttaattaa attagtTTT	4680
ttatTTata ttGAAGAAA ttGAAGTTT ggaattggag agtattatgt taaatgaaat	4740
aagttaaata tagaaagata aatattataat gttttattt atttgtgaaa tataaaataa	4800
ttatTTTTT agtagtaaag agtagaatgg tggttattag agttgggggg tgggaggaat	4860
ggggagatgg taattaagat ataaagtTTT agttaagatg ggaggataaa gtttgattgt	4920
ttttttgag atgtgtttt tagtatgtg aatatagtt aatagtaat ttAAatgtt	4980
tttatttgat aaaaatgtt aatatttgat atgatggata ggttatttag ttgattttaa	5040
taatTTTTA ttgtgtttaa agattataat ttatattgtt attatataaa tatataataat	5100
tgtattattt taatatataa ttTAaaaattt aatataatga aaaAGAAattt gaagttaat	5160
atTTTTtagaa gttaagtgtta attaaaagt ttgtgagaa ttgttttaa taaataaata	5220
agttttttt tttaataat tattatTTT tttgtttggaa tatatagttag tgaataaaaa	5280
aaaaaaaaaaa aaaaaaaattt tttaggttta atataattt aggaagaaat tttagtagtt	5340
gtatTTTTT ggaatataat gaagttatgt tggtgtttt gtttagttgt ttttggTTT	5400
ttgttatttt gtttggttt tatagtgtt tttgtttgtg atgatagtTT tttgttttt	5460
tggaggatAT aatggaaattt attgtgtattt gaagaatggaa tagagaattt aagaaggaaa	5520
ttggaaatttT gaagtaatgt tagggtaat tagatattt gggTTTgtg ggggtttgt	5580
ttggTggTga ggggtttta tataagtTTT ttTTTgtta tggtgtttt tattttggTT	5640
ttgattattt tgTTTTTTT ggttagg	5666

```

<210> SEQ ID NO 20
<211> LENGTH: 5666
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

```

<400> SEQUENCE: 20
ttgttagag agaataGAGt ggtagagtT agggTggggg ttgtatgtg ggaaAGGAAG 60
ttgtgtaga gttttttat ttttaagttag attttatAT aagtTTtagg tgTTtaattt 120

```

-continued

tttttatatt ttttttagt tttaatttt ttttttgagt tttttattta ttttttagta	180
tataatgaat ttatttatat ttttgaatt ttgtggagt ttttgtata ggtagagagt	240
attgtgaggt atgggtaaaa tagtaaaggg gtagggatag attgattttt atttttaggtt	300
aatttttgtt atttttttt agatataattt attgaaattt tttttgaaa ttatgttagg	360
tttggagatt tttttttttt ttttttttt tgtttattgt tttatattta agtgtagaat	420
gtggtaattt tttttttttt ttttttttt tgtttattgt tttatattta agtgtagaat	480
ttaagttata tttagttttt gggaatgttg aattttaaattt ttttttttattt tatatttagtt	540
ttaaaaattt atattggat agtatagttg tatataattt tttgttttataa tatgaagttt	600
tgatttttga atataatggg gaatttattaa gttaagttaa gtaatttttattt ttttttttta	660
aatatttgat atttttgttta aatgagagta ttgggattt attatttagt tatatttttattt	720
atgttatgaa atatatttttta aaaaaaataaa tttaatttttattt ttaatttgagg	780
ttttatattt tgatttattt tttttttttt ttagtttttag taatttttattt	840
tttattttttt attgttaaga atgttaattttt atagataatgt gagaatatgt	900
gatattttgtt tttttgtt tggtttttttt tttttttttt aattttttttt aattttttttt	960
tttaatttt tttaagttata aaataagaag gtttagtttaa ttaatttttaa aattttttttt	1020
tgtggtaggt tgaataatgt tttttttttt ttaatgtttta tttttttttt tttaaaaaattt	1080
tttaatataat taatttatgt ggtttttttt gttttttttt tttttttttt ttaatgtttt	1140
tgagggagat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1200
gagggttata gtttagagaga agatataaga atggaagtat aggttataaga gaaaatataag	1260
agattatgag ttaaggaatt tttttttttt tttttttttt tttttttttt tttttttttt	1320
gtttttttaga gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1380
tagaattttgttta aaataataat tttttttttt tttttttttt tttttttttt tttttttttt	1440
atagtagtag taggaaatttta aaatattttt taggtttatg atttgagagttt tttttttttt	1500
agagatggttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1560
aattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1620
gttagataattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1680
aaagaggagt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1740
taaattaagg aatttatatta aatttatatta tttttttttt tttttttttt tttttttttt	1800
atttgattttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1860
aaaagaaaaga tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1920
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1980
aaagtaattt atatgttaat tttttttttt tttttttttt tttttttttt tttttttttt	2040
attaattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2100
atataattttt atgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2160
tattaaatgttta gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2220
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2280
ataaaggttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2340
tggtaaaattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2400

-continued

attttttat	aaaataggag	tagggagagg	agaaatattt	ttgtttaaaa	atgaggaatt	2460
gaaaattttt	attataaata	aattatatta	agtaagttaa	agatagtaaa	agagtaaaaa	2520
tgttagtaga	tattttaaa	atggtaatta	tatattattt	ttggaatgat	tatatgaatg	2580
tggtttatta	tttttaagt	tttatagta	aatatatatt	tatgtttttt	attttagttaa	2640
aaataaataat	aatatgtagt	tgttttgaa	taatttttt	ttttttttt	ttttttttt	2700
ttttttgat	aaagttttat	tttggttattt	aggttggagt	gaagtggttt	tatgtttgtt	2760
tttattataa	tttagtttt	ttgggtttaa	gtgattttt	tgttttaatt	tttgagtag	2820
ttgggattat	aggtgtttgt	tattttttt	ggttattttt	tgtattttta	gtagaggtga	2880
ggttttattt	gttggtagg	ttgggtttga	attttgattt	ttaggtgatt	ttttttgttt	2940
tgatttttta	aagtgaaggg	attataaggt	gtgaggtt	gtgtttgggtt	gtttttgaat	3000
aattttgatt	aaaatttata	tttgataattt	attttaatat	atattataga	tttttattga	3060
taatttttt	tagtaagaaa	gataagttt	attttaggtt	ttgtgaattt	gaggtaagt	3120
agtttttagta	tattttataat	tttttaaga	ttttttttt	attttaatg	tttgtaaatt	3180
ttgtatttga	taaagagttat	atttttattt	aatataaaata	tgtttttttt	tttagatttt	3240
tttagtattt	gagagattt	tatgtgtgt	tttttttatt	ttttttttt	gtttttttaa	3300
tttttttaggg	ttgtgttagg	aggaggttt	tgattataaa	tttttttga	aaattttta	3360
ggaagttttt	ttttttttt	gagaattgaa	gtgttattt	attttaattt	tttgtaaat	3420
tttgtttttt	agagttgttt	tttatttttt	ttttttgtt	tagatttttt	atttatttgg	3480
atgggaaaa	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3540
tgtatttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3600
atgttttttt	ttgtgtgtt	ttttatgtt	ttttttttt	ttttttttt	ttttttttt	3660
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3720
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3780
ggagggattt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3840
gttattaatt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3900
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	3960
atgggaaaa	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4020
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4080
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4140
tgtttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4200
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4260
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4320
tagtttaag	gatttttttt	ggagagttt	attttttttt	ttttttttt	ttttttttt	4380
tttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4440
aagagtagat	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4500
aggttagatag	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4560
aatttttaga	ttttttttt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4620
gattagtttt	aaatattgtt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	4680

-continued

aaatattttt tatttttgtt aattataaag ggggaaggga atatgagtgt tttttgtttt
ataggggttt ttgtgagtt aatgtatgt ttaatatata taagtttaa gaatagtgtt
atatattta agttaatatt tgtagtttt tgaattattt gttttgagga ttgggttgta
atttgtttt gaggtataga aagaaaatgt tttggagtag gatgtggtgg tttatattt
taatttttagt attttggaa gttgaggtgg gtagattatt tgaggttagg agtttgggt
tagtttgggtt aaaaatggta tattttgtt ttataaaaaa tataaaaatt agttggttat
ggtgggtgtat gtgtgttaatt ttagttttt aggaggttaa ggttaggagaa ttgtttaat
ttgggaggta gaggtttagt taagttgaga ttgtgttatt attttttagt ttgggtgata
aatgagatt ttgatttaaa aaaaaaaaaa aatgtttgg atagaattat tattattata
taaaaggaaa gtttggatgt ggtggtttat gtttataatt ttagtatttt gggaggttga
gataggtgaa ttatttgagg ttaggagttt gagataagtt tgattaatat ggtgaaattt
tgtttttatt aaaaaatata aaatttagtgg ggtttgggg tttatgtttt taatttttagt
tattttggagg ttgtatgttagg agaattgttt gaatttagga gaaggtggag gttgtagtga
gttgagattt tgttattgtt ttttagttt ggagataaga gtgaaaattt gttttaaagaa
aaaaagaaaag aaagaaagaa agaaagat tttttttt aagattatgg
gtatatttttta ttatatttttta ttatataaagaa aagtttaataa gtatataaaga gtataataag
tcttaaggaaq taaaagttt aatttt
5666

```
<210> SEQ ID NO 21
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

cgcggatccg attttaatgc 20

```
<210> SEQ ID NO 22
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

<400> SEQUENCE: 22

<210> SEQ ID NO 23
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 23

cgacgaaatt cctaacgcaa ccgcttaa
<210> SEQ ID NO 24
<211> LENGTH: 20
<212> TYPE: DNA
<212> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 24

ttcggatgg gaacgggtgta

20

<210> SEQ ID NO 25
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 25

ctccccaccgc cgttacc

17

<210> SEQ ID NO 26
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 26

cccgtcctaa ccgtccgcc t

21

<210> SEQ ID NO 27
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 27

tcgtcgttgt ttccggtagt t

21

<210> SEQ ID NO 28
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 28

ccctccgaaa cgctatcga

19

<210> SEQ ID NO 29
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 29

cgaccataaaa cgccaacgcc g

21

<210> SEQ ID NO 30
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 30

-continued

ttttttttc ggacgtcggtt g 21

<210> SEQ ID NO 31
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 31

cctctacata cgccgcgaat 20

<210> SEQ ID NO 32
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 32

aattaccgaa aacatcgacc ga 22

<210> SEQ ID NO 33
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 33

tggaattttc ggttgattgg tt 22

<210> SEQ ID NO 34
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 34

aacaacgtcc gcacccct 19

<210> SEQ ID NO 35
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 35

acccgacccc gaaccgcg 18

<210> SEQ ID NO 36
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 36

gaaccaaaaac gtcggccat 19

-continued

```
<210> SEQ ID NO 37
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 37
```

```
tatatatgtcg gttacgtgcg tttatata 27
```

```
<210> SEQ ID NO 38
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 38
```

```
cccgtcgaaa acccgccgat ta 22
```

```
<210> SEQ ID NO 39
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 39
```

```
gcgtcggagg ttaagggttgt 21
```

```
<210> SEQ ID NO 40
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 40
```

```
ctctccaaaa ttaccgtacg cg 22
```

```
<210> SEQ ID NO 41
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 41
```

```
aactcgtcg cccgcccga 19
```

```
<210> SEQ ID NO 42
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 42
```

```
tttcggatgg gaacgggtgt 20
```

```
<210> SEQ ID NO 43
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
```

-continued

```

<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 43

ctcccaccgc cgttacc                                17

<210> SEQ_ID NO 44
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 44

cccgtcctaa ccgtccgccc t                                21

<210> SEQ_ID NO 45
<211> LENGTH: 2501
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 45

cttggactct aatgtgtatt ttacacttac agcacaatta atttggact agctacattt      60
cagctcaaca atagccaata gcatatggga tagcgcaaat aaactctgcg tctctgtgc      120
ttctttgggt ctcggagacc tcaacccttt cttcagatg caaaccttct tgcccttcaag      180
cctcggtctcc aacaccaggc cggcagagga acccagtcta atgaggtacg ctcccttcct      240
gccattctctt attccattaa cctgttttgt ggttaaacgtt ggactgtatcc tccaaaatta      300
ccttattaaat tagcttacat atttattttt tatctgtccc accagaatgc aggtttccgg      360
aaggcaggga tttaaaaaaa tctgttttgt tctatgtat tttccctatac caagcaccgt      420
gccccggcaca agctgggatc ccagttacaca ttcggggacg gaagaaccgt gtttcccttag      480
aaccctaggta gagggcagct tagcaatgtg tcacagggtgg ggcggcccg gttccggccgg      540
acgcactggc tccccggccg gctgtgggtgt ggggcgagtg ggtgtgtgcg ggggtgtgcg      600
ggttagagcgc gccagcgacg cgggagcgcg gagctgggag gaggcagcgcg cgccgcgcag      660
aaccctggcgc gccccggctgg cagggcagct cggagggtgg tggggcccgcc cgccagcccg      720
cttgcagggtt cccccattggc cgcctggccg cggccctcccg cccaaaaggc ggcaaggagc      780
cgagaggctg ctccggagtg tgaggaggac agccggaccc agccaacgcc ggggactttg      840
ttccctccgc ggaggggact cggcaactcg cagcggcagg gtctggggcc ggcgcctggg      900
agggatctgc gccccccact cactccctag ctgtgttccc gccgcggccc cggttagtct      960
ccggcgctgg cgcctatggt cggcctccga cagcgctccg gagggacccg gggagctccc      1020
agggcgcccg gtgagtagcc aggcgcggct ccccggtccc cccgaccccc ggcgecagct      1080
tttgctttcc cagccaggcc ggggtgggt ttgtccgggc agtgcctcg gcaactggg     1140
aggecaaggc ggagggaaac ttggcttcgg ggagaagtgc gatgcagcc gggaggctc     1200
cccagcccg cggggccgggt gagaacaggc ggccgcggcc cgaccaggcg ctttgcgtcg     1260
gggcgcgagg atctggagcg aactgctgcg ctcgggtgg cgcgtccctt ccctccctt     1320
ctccccccgg cggccgcacg cggggcggc cgggtAACGG agaggaggtc gccaggaaatg     1380
tggctctggg gactgcctcg ctccggaaag gggaggggt ggccacgggt ttaggagagg     1440
cgccggagcc gagagggtggc gccccgggtgc caccgttgcc gcaggctgga gagagattgc     1500

```

-continued

tcccagttag	gcccgttaccc	tctggggcag	ggcttcattc	ttccgcggcg	ccccctggagg	1560
tggaaagct	gggtgggcat	gtgtgcagag	aaaggggagg	cggggaggcc	agtcaactcc	1620
ggagccggtt	ctgatcccaa	cagaccgccc	agcgtttgg	gacgcccggacc	tcggggtgcc	1680
gtggtgcccg	gccccacgcg	cgcgcggggc	tgaggggtcg	ggggcgcccc	tggccgcccc	1740
gcttaacaa	agggtgtcc	tctccacccc	gcgaggagg	gcagctccgg	agaccgggtc	1800
ttcagcgagc	ggggcttag	cgcgcggggag	gtctacttcc	ttttggggtt	gccattttac	1860
tattattatt	gcctttttt	tttcttcaaa	aggactggag	actgatgcat	gagggggcta	1920
cggaggcgca	ggagcgggtgg	tgtgggtctg	ggaagcggag	ctgaagtggcc	ctggggtttg	1980
gtgaggcggt	acagtttatac	atgaccgtgt	tcaggcagga	aaacgtggat	gattactacg	2040
acaccggcga	ggaacttggc	agttaaagg	ggtaccagaa	gcgtaccctc	ctggattgtg	2100
gaaatgcata	acgatggggc	cattgggtgg	taaacaaaatg	cagtttgaat	caggcgctc	2160
cctcgccctt	tctggagatg	cgcaaatcat	agagaaaaga	gttactaacc	cagcggtaaa	2220
cgcctgtatc	caagggcctg	gggggtggagg	agaggcagca	gttcagggt	agattatgat	2280
gcacagtata	ttgatccagt	cccctggaca	aaatcagat	taattgtccg	tgctaaactct	2340
tgtcagccct	tgccttctg	tgacaacagg	acaaacacta	agattataat	tgcaatttgg	2400
gttagcttt	atgtgtgatt	taaacggagg	gtacaaacta	attaataggt	tttaaaaatc	2460
ttagtacttt	accctctatc	taaattttca	gtgttaatttgc	a		2501

<210> SEQ_ID NO 46
<211> LENGTH: 4501
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 46

ttcacttgtc	ctacaggatt	ccccatggaa	tcttggagtt	tttgaggcga	gagggtatcct	60
ggataaccact	gagttctatc	tttcatccaa	taaacacaga	agtggacgcc	tggacaggca	120
aaegtacttg	accaaggcag	gtgcacagct	attctgcaac	attggaaaca	aatctcaggt	180
cttttgcattt	tttgcatttca	ctttactctc	ttttcatttc	ccagaaacaa	agttttcatg	240
tgcttttttt	tatagtgata	tgtttggat	gcattagcta	gtatatttgg	aaggaaaaaa	300
aataaacaca	caagagataa	acctgtcagg	aggacaaacc	tgtattgct	ctgattggct	360
cagagggtga	ttatttatcat	ggttagagaat	tatthaatca	gtgtaaatgg	aattttctcg	420
tgggctgggc	actgtacaaa	gactcaaacc	aatctgtcta	cagatctgaa	aagcagatac	480
gagatctgtg	aatggctggg	gtttccaagg	ccacagtaca	agcatgggcc	acaccttaca	540
gcttggagga	ctgagccctg	aaaatggca	agttccttca	cttctctgaa	ccttattttt	600
ccccacattt	aaacaaggat	gagtagttt	tgagggtcctt	tttacgactt	cttctctac	660
agactcttagc	atccataac	ttgataaaaa	gagggtggat	atgaactcac	cttctctaga	720
aaagttccag	gaaagagaat	accaggtcat	ccttagtaggt	gtgtagacag	gccagataga	780
tcttgcact	tactcagttc	ttcccaagatg	tataactcta	tcattgttct	tagctgtcaa	840
gagaaagcag	gagagcctgc	atcttcattc	ttttttttt	ttttttttt	tttggagacg	900
gagtctctact	ccatcaccta	ggctagagt	cagtgccatg	atctcagctc	actgcaagct	960
cgcctccca	ggttcacgccc	attctcctgc	ctcagcctcc	caagtaactg	ggactacagg	1020

-continued

cggccaccac cacacctggc	taattttttgc	tgttgttagt acagacgggg	tttcaccatg	1080
ttagccagga tggtctcgat	ctcctgacct cgtgatccgc	ccaccttggc	ctctcaaagt	1140
gctgggatta caggcgtgag	ccaccgcacc cagcctgcat	cttcattctt	actgttagcc	1200
tcaggttcac cccacctagc	ttattaagt atgttgaata	accaattctt	acatattatt	1260
aggctcatgg acaccatgac	atccagactg atgggtgcct	gctgaagggg	gtgacccatg	1320
caggaggact cccctacgca	aggattcatg gagtttgctg	tttctttcc	ttagggtgag	1380
aaccaaactg cttcacacg	gtgggcagag gggaaactgac	tcaggtttg	aataagagag	1440
aacatccaa ctgaaaagct	cttggattc gctgaacttc	aagacactgt	gtggacccagc	1500
ttaggatagg gagtgagaag	aaattaacca aaaggtaatt	tcgttacttt	tcaagctggaa	1560
aaaagatcag attatacttg	tgctttcata attaagttagc	tgctggaaaa	aaacgcttca	1620
gatgcttct atgagaaaac	tgctgcttga agttcagcag	aagttatcta	cttgatactt	1680
atattccagg caaggccttc	cgttggagaa aatatcgca	ctttggacaa	aactgaaatg	1740
tgaaaagaaa gggaaagagag	ggcctctatac atgtaagatg	cttacccaaa	gtggatttg	1800
tctggaaagt cttctaaaac	cttccacatg actgtggaat	aagtcatgtg	gggcgcgggg	1860
ataagcgaat ctctcaaatt	ccaccacgta tgccctcatt	caacctggat	ccttagagtg	1920
gcctccaggg cactctgttc	aggactcagt cagctgttgg	ccacaccat	gctctcagtg	1980
ctcctgagac cctatttgg	tctgagaggg ctaaaaagca	gtgtggctaa	atatcccagg	2040
cctcaaagta ttctactgt	ggttggggaa gcaatagaat	cataccccat	aaaacaatga	2100
aaacagtgtc agaaaaacat	cgagagacag aaacatctc	acgagttagg	ccacagttag	2160
agtgaaggca gggaaaggttt	ttaaagctgg gtggagggga	caagtcaaaa	agatgtggaa	2220
actggttcc cttcctatg	gctaaagtgc tcaaagggg	aaaaggagtt	tcaaaaatgt	2280
tcttgaaat accatcttc acgaattctt	cgccctctgc tgcctcaatg	tcacttgtct	2340	
gagatgtaaa cagaggagtt	ctgagaaaaga agctgaactt	gcatttctcc	ctgtttctat	2400
ttgttccaaa cttgtggcat	ttctaacagg atgaagcgga	agagaaaagg	aaagagacaa	2460
aagtgtagaa agatggaaga	teccagctgc aaatggccat	ttgcagttag	atggaacacgc	2520
tgctgacgtt cagggaaatg	catgtctctc ttcaagatgg	aaggagcagt	ggaaagggg	2580
gacgagttcc tggctggcca	ccaatcatcc catcttctg	tgcgggttcc	tcatctggaa	2640
agtggagtg atacttgtc	ttgttttcc tacccacaaa	gattattgtg	agagctataa	2700
tacgggtgaga tacagaatcc	tgcttttaaa aatacaaaagc	agaatcaaga	tgtcaataat	2760
aaggatagta attgtgttag	ttatctgcata tcatacttata	tagctagtc	tcttagatcc	2820
tggatcggtc tcctgggttt	actacagttt tggatcagct	caccccaaa	tccctgtcg	2880
aagggtggag ctctgtcagc	catggcagg gaaccactc	ctcttgcctt	tctactttct	2940
gtctttcaaa catgccagg	gtctttgcac ttgctgttcc	ccctgectgg	tacctctctc	3000
ctgtggctt ccccaagatgt	gatcctgtc ttgttccact	tctcaagcgag	gatggcactt	3060
cagggagccc ttcccttact	atcgcaagaga gagcaggccc	tccccagtc	tgtccaaccc	3120
agaactctgt ttgtttct	tcatagccct agcatcacag	aaaatcaccc	tgtgcattca	3180
tggatgtcca cggggcaag	ggctttgtgt tgcttaaccc	agcatcctga	accgtgtttg	3240
ttgaatgaat acagaacccc	gtttgctctg ggagagcaca	gaaaacagtc	ttctatcata	3300

-continued

tatcatagcc agctgcaaac	agaaccagag	3360
agcagatggc ttcccata	ccagagaga	
tcatacatac tttgggtctt	tccctctgt	3420
ctggagggtt agcagttagc	attccccgc	
tggttccacc aagcacatc	aaggctct	3480
ggacatggcc acccctcacc	tgtggaaagcg	
gtcctgttgg ggtgggtgg	tgtagttgg	3540
ttctggttt ggtcagagac	acccagtggc	
ccagggtggc gtggggccag	ggcgcagac	3600
agaaggggca cgagggtctc	gtcccgagg	
cccagggca acccctcacc	ccacccggca	3660
ccacactcggt gcacccggc	gcattattcc	
ctataaggat ctgaacatc	cggggccggc	3720
ccccccgt tacccttgc	ccccggcccc	
gccccctttt tggaggggcc	3780	
atgaggtaat gcggtctgc	cattggtctg	
agggggcggg ccccaacac	ccgaggcg	3840
gtccccgggg gcccagcg	atatactcg	
gcccggccagg cagcggcg	gagcggcg	3900
caggcaggcg gcggggcg	agacggctc	
tccttcctctt cttgtcc	cagtcctgc	3960
tccttcgccc ggaggccg	cgccgagtcc	
tgccggcagcg ccgaggc	ctcgctgc	4020
ccccccgtt ccggccggc	actcggagg	
cagcggcgcc gaggccaagg	ttgccccca	4080
cggccggcg	ggcgagcg	
ctcggtctgc agcagcccg	ccggccggcg	4140
tttggagagg cgagcagc	ccccggc	
ggggcagca gggcaatga	ccccttg	4200
cgggctcatc tggtctctgg	gcagctgg	
cctggggac tggggcg	aggcgtgc	4260
atgctcgccc agccaccc	aggacgc	
ctgcaactcc gacatcg	agcgctct	4320
gtccccggcc cgagcccc	gctgcagcc	
ggactgc	gctgtttag	4380
gagccccact ctttcctt	gccccagg	
agggcagac ggggttggg	cgagatgg	4440
aaactcgatg tcattgg	ggggcg	
ctggcagtg	catagctg	4500
g		4501

<210> SEQ ID NO 47

<211> LENGTH: 3001

<212> TYPE: DNA

<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 47

gaagtgtctaa tgcgttgc	ttacccacta	cataagccca	ctcttgtact	agggcagtga	60
ctttcttctt tgggtgagac	cttgcattct	gggattataa	ttttgaatta	taattataaa	120
atggtatttg gctgtaaatt	atctcccttt	tttttctgtt	cctcacat	gatattatgg	180
attccatataa ggattcatgt	cttcttattca	ctttaatgaa	cagttgttgg	gcaacaattc	240
tagaagagtt ccaattctca	tcaggagaat	ggacaagg	gagaagcaga	gaaaatgcaa	300
ttagttagaat gcttaagtca	tcacttttgg	attgactgaa	cataaataaa	aatgagaaag	360
atacgtaaaa aagaaggaa	tggtaagca	gggtgtatgc	tgggagagg	ggggctccat	420
agccatgaga gtcaactctg	taacaccctt	tagggttaca	acactgc	tcataactg	480
aggtagcagc agggaaactt	ttaattatt	agaaatattg	aactttgc	cccacccca	540
aacatttttc tcattcagtt	cctgttctt	tttatttctg	taattttac	tgttcaaaa	600
atgatctttt ttcttcgga	agaagcaatt	cttcaaatcc	agttcacata	aggggattt	660
atatgttcaa caagctccaa	atacactgta	tccagcaata	cctactacat	gcctacttt	720
agctctgagc aacctgcacc	tcaagcctag	ttctcattgt	tttgcatttt	gcaaatttc	780

-continued

actaagtgcc	cttccccc	aaacacacgt	atatgtctac	cagaccctaa	agcccttat	840
gaacatcaa	actcctccct	tctgaaaacc	tttgcgtgag	tggtcagcag	gctaattcat	900
ccattgcaat	gtggcttgt	gttagggttc	tgtttccgtg	ctgcctgcaa	gataatcaca	960
gatgtgactg	catcttagaa	gttcctgaat	ctttcaagac	agtctggttc	acaagaaaat	1020
taaaagggtgg	aggtcggcg	cggtggctca	cgcctgcaat	cccagcactt	tgggaggccg	1080
aggcgggcgg	atcacctgag	gttgggagt	cgaaaccagc	ctgaccaaca	tggggaaacc	1140
ccgtctctgc	taaaaataca	aaattagcca	ggcgtggtgg	tgcatgcctg	taatcccagc	1200
tactcgggag	gctgaggcag	gagaatcgct	tgaaccggg	aggcagaggt	tgcgatgagc	1260
cgagatcggt	ccattgcaet	ccagcctggg	caacaagagc	gaaactctgc	cacacacaca	1320
caaacacaca	cacacacaca	cacacggtgt	agtttagaa	gtaaaaaaaaa	aaaaaaaaaa	1380
aaaatcagat	ctcccctcac	acctcagatc	tgaaggcaca	aactctaggg	ccagggcggt	1440
cgcctaccca	actccacatg	cacttgcagg	tcacctagca	ctcaggtacc	tagcactcag	1500
gtacattgtg	gtccttacc	tctcacgaca	gcagcaacaa	cgttgatgg	aagtttatca	1560
ctgtgtgtta	cggccatgg	gccccatgg	gccccatgg	ttagaatttt	atgtgaaatt	1620
tctcacggac	acccctgaaa	cagatgccac	agccccatt	ttgccaacga	ggcagctgag	1680
gttcccagag	gctcaatacc	agcaccatga	gcccgcac	gcaaggcataa	cacagccgga	1740
ggtgagcaca	tacctgcttc	gcacccatg	cgccctaacca	caaggttccc	tccctccagg	1800
aaggccgttg	tctccctgg	gacgacttgc	cagctctgag	gcatgacagt	acggggcccc	1860
agaagggtga	ccaggaggcc	ctoctcgccc	cagctgcgg	cgtcgcgc	cactgoaggg	1920
ccggggctgt	gactcggtgg	gacggttccc	tgcgccccgg	cggggggggt	gggcggggag	1980
ggggggccgg	gccccggggc	ggggctcggg	acggccgggc	tgggagctgg	agccccacagc	2040
gggaagccgc	cgccgccegg	gctcgcagg	gttagggcag	gccccggggg	gcggggccgg	2100
gcccgtacggg	aaggggaggc	cgccgcggacc	gggagccgca	ccgcgcac	cgggctgcag	2160
cgccgcgcga	ccaaggctgc	gatggggctg	gagacggaga	aggccggacgt	acagctttc	2220
atggacgacg	actctctacag	ccaccacagc	ggcctcgagt	acgcccaccc	cgagaagttc	2280
gcccgtcgg	accaggaccc	ggatccccac	cggtctcaact	cgcatctca	ggtgaagccc	2340
ggggccggcgg	ggcccaagtc	cccgctgagg	ccggggagggt	cgccgcgc	tcageccccgc	2400
cctaaccctgt	cccaccattg	ctaccgggtc	ggcccccgcag	ggtctgagac	ccgcaccctt	2460
ccccggctcc	acccgtcacc	aggccgcgc	cgtagccagg	aattcttac	caggttctgt	2520
tgcgcgcacc	gtgaccctaa	gagaagaggc	ggacgcctg	gcacgtcctt	ccctctgtct	2580
tccccggccc	aaagcgctcc	cggttcccg	ggcgtcaggt	tggctgacag	ttcggggtcc	2640
ctgegtcttg	tctcctcagc	tggcttcga	ggatgtgatc	gcagacgg	tgactacgc	2700
ctcctttgac	aaagtgtgga	tctgcagcca	tgcctcttt	gaaatcagca	aatacgtaat	2760
gtacaagttc	ctgacggtgt	tctggccat	tccctggcc	ttcattgcgg	gaattcttt	2820
tgccaccctc	agctgtctgc	acatctggtg	agacggggca	caccgggtgg	accggcttcc	2880
tgaaacatgg	gcatattctc	cgccacactgc	cccctactct	cctcttatcc	caggccggcg	2940
tcaggaggag	gaacgcgcac	cagttcccaa	cgagtaggaa	gaactgaaag	gccttgaaag	3000
g						3001

-continued

```
<210> SEQ ID NO 48
<211> LENGTH: 2501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 48

tttggatttt aatgtgtatt ttatattttat agtataatta atttgggatt agttatattt      60
tagtttaata atagttataa gtatatggga tagcgtaaat aaattttgcg tttttgtgt      120
tttttgggt ttcggagatt ttaatttttt ttttagatg taaatttttt tgtttttaag      180
tttcggtttt aatatttagtt cggtagagga atttagttt atgaggtacg ttttttttt      240
gttattttttt attttattaa tttgtttcgt ggttaaacgta ggattgttattt tttaaaaatta      300
tttatttaat tagtttatattt atttattttt tatttgtttt attagaatgt aggttttcgg      360
aaggtaggga tttaaaaaaaa tttgttttgcgt tttatgtgtat ttttttatattt taagtatcgt      420
gttcggtata agttgggatt ttagtatata tttcgggacg gaagaatcgt gtttttttag      480
aatttagtta gagggtagtt tagtaatgtt ttataggtgg ggcgttcgcg tttcgggcgg      540
acgtatttgtt ttttcggcgt gctgtgggtgt ggggcgagtg ggtgtgtgcg ggggtgtgcgc      600
ggtagagcgc gttagcgcgt tccggagcgcg gagttgggag gagtagcgcg cgtcgcgttag      660
aattcgttagc gtcgggttgg tagggtagtt cggaggtggg tgggtcgcgt cgtagttcgt      720
ttttaggggt ttttattttgtt cgtttgtcg tctgttttcg tttaaaaaggc ggtaaggagt      780
cgagagggtt tttcggagtg tgaggaggat agtcggatcg agttaacgtc ggggattttg      840
ttttttcgc ggaggggatt cggtaattcg tagcggtagg gtttggggtc ggcgtttggg      900
agggatttcg gttttttattt tatttttagt ttgtgttttgcgtt cggtagtttt      960
tcggcgttgg cgtttatgtt cgggtttcga tagcgtttcg gagggatcgg gggagtttt      1020
aggcgttcgg gtgagtagtt aggcgcgggtt tttcggtttt ttcgattttc ggcgttagtt      1080
tttggttttt tagttagggc ggggtgggtt ttgttcgggtt agtgtttcgta gtaattggga      1140
aggtaaggc ggagggaaat ttgggttcgg ggagaagtgc gatcgttagtc gggaggttt      1200
tttagtttcg cgggtcgggtt gagaatagggtt ggcgtcgggtt cgattaggcgtt ttttgcgt      1260
gggegcgaggg atttggagcg aattgttgcg tttcggtggg tcgtttttt ttttttttgcgt      1320
tttttcggg cggtcgtacg tccggtcgggtt cgggttaacgg agagggagtc gtttaggaatg      1380
ttgttttggg gattgttgcg ttcggggaaag gggaggggtt ggttaacgttgcgtt ttaggagagg      1440
cgccggagtc gagaggtggc ggggggggtt tttcgttgcgtt gtaggttggc gagagattgt      1500
tttttagtgcg ggcgtatcg tttggggcgtt ggtttttttt tttcggggcgtt ttttggggcgtt      1560
tgggaaagtt ggggtgggtat gtgtgttagag aaagggggagg cggggaggggtt agttatttc      1620
ggagtccgtt ttgattttaa tagatcgatc tttcgttgcgtt gacgtcgattt tttcggggcgtt      1680
gtgggtttcg gttttacgcg cggcgccgggtt tgaggggtcg ggggggtttt tttcggggcgtt      1740
gttttaataa agggtgtttt ttttttttttgcgaggagg gtagttcgg agattcggtt      1800
tttagcgcgc ggggttttag cgtcggggag gttttttttt ttttgggggtt gtttttttttgcgtt      1860
tattattattt gttttttttt ttttttttttgcgaggagg attgtatgtat gagggggtt      1920
cgaggcgttca ggagcgggtgg ttagtgggttgcgtt ggaagcggag tttcggggcgtt ttttggggcgtt      1980
```

-continued

gtgaggcgtg atatgttttatt atgatcgtgt ttaggttagga aaacgtggat gattattacg	2040
atatcggcga ggaatttgggt aggtaaaggg ggtatttagaa gcgtattttt ttggattgtg	2100
gaaatgtata acgatgggg tattgggtgg taaataaaatg tagtttaat taggcgttt	2160
tttcggtttt tttggagatg cgtaaattat agagaaaaga gttattaatt tagcggtaaa	2220
tcgtttgatt taagggtttg ggggtggagg agaggttagta gtttagggtt agattatgat	2280
gtatagtata ttgatttatg tttttggata aaatttagatt taattgttcg tggtaatttt	2340
tgttagttt tgtttttttg tgataatagg ataaatatta agattataat tgtaattgga	2400
gttagtttt atgtgtgatt taaacggagg gtataaaat attaataatg tttaaaaattt	2460
ttagtattt attttttatt taaattttta gtgttaattt a	2501

<210> SEQ ID NO 49
<211> LENGTH: 2501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 49	
ttaaattata ttgaaaattt agatagaggg taaagtatta agattttaa aatttattaa	60
ttagttgtta ttttcgttt aaattatata taaaagttaa ttttaattgt aattataatt	120
ttagtgtttt ttttgttggt atagaagggt aagggttgat aagagtttagt acggataatt	180
aaatttgatt ttgttaggg gattggatta atatattgtg tattataatt tagtttgaa	240
ttgttgtttt ttttttattt ttaggttttt ggattaggcg gtttatcggtt gggtagtaa	300
ttttttttt tatgatttgc gtattttagt aaaggccgag ggagacgttt gatttaattt	360
gtatttgtttt attatataat ggtttatcg ttatgtattt ttataatttta ggagggtacg	420
tttttgttat ttttttattt tggtaatgtt ttgcgtgggt tcgttagata tatttacgtt	480
tttttgtttt aatacggta tgataaattt ttacgttttta ttggtaatttta gggtagtttta	540
gtttcggtttt ttagattattt attatcgttt ttgcgttttc gtagttttt tatgtttag	600
tttttagttt ttttgaagaa aaaaaaaagg taataataat agtaaaatgg taattttaaa	660
aggaagtaga tttttccgc gttaaagattt cgttcgttga agatcggtt ttccggatgt	720
tttttttccg cgggggtggag aggagtattt tttgttaaag ttggggcggtt agggacgttt	780
tccattttttccg attttcgcgc ggcgtgggg tcgggtttaa cggatattcg aggtcgccgt	840
ttttaaacgt tggccgggtt gttgggattt gaatcggtt cggaaagtgtat tggtttttc	900
ttttttttt tttttgtata tatgtttattt tagttttttt atttttaggg acgtcgccga	960
agaatgaagt ttccgtttttag acggtaegcg ttttattttttt agtaattttt ttttagttt	1020
cggtaacgggtt ggtatccgtt ccgttattttt cgggtttccgc gttttttta atatcggtt	1080
tattttttttt ttttttccga cggaggtattt ttttaggtttt atatttttgg cgattttttt	1140
tccgtttatcc ggtcgatccgt cggatccgggtt gttccggggtaa gtaaggaggaa gaaggagcg	1200
gtttatcgat gctgtttttagt tcgttttgc ttttcgtttt tcgtatataaa gcttttgcgtt	1260
gggtcgccgt tattttttttt tattcggttcc ggggggtttt ggaagttttt cgggttgcgtt	1320
cgtttttttt ttcgaagttt agttttttt cgtttttttt tttttttttt ttcgaggtat	1380
tgttcggata aattttatcg cgtttttttt gggaaagttaa aagttggcgtt cgggggtcggtt	1440

-continued

ggggatcggg gagtcgcgtt tggttattta ttccggcggtt tggaggtttt ttccgttttt	1500
tccggacgtt gtcggaggctc gattataggc gtttagcgtcg gagattagtc gggcgccgg	1560
cggaaataata gtttagggagt gagtgggggg cgtagatttt ttttaggcgt cggttttaga	1620
ttttgtcggtt gcgagttgtc gagttttttt cgccggaggga ataaagtttt cggcgttggt	1680
tcggttcggt tggtttttt atatttcgaa gtagtttttc gggtttttgt cggtttttgg	1740
gcggaggcgc gtcggtaggc ggttaatggg gatTTTgtaa gcgggttggc ggccgggttt	1800
atttatTTTC gagttgtttt gtttaggtcg gtttgcgggt tttgcgcggc gttcgttggt	1860
tttttagtt tcgcgtttcg ggttcgttgg cgcgTTTtat cgcgatatt tcgtatatat	1920
ttatTCgtt tatatttacg tcggtcgggg agtttagtgcg ttgcgttgcg acgcggcgt	1980
tttatttGtg atatattgtt aagttgtttt ttgattgggt tttagggaaa tacggTTTT	2040
tcgtttcag atgtgtattt ggatTTTtagt ttgtgtcggt tacgggtttt ggtatggaa	2100
aattatataag aataaaatag atTTTTTaa atTTTTgtt ttccggaaatt tggatTTTgg	2160
tggatagat agataataaa tatgttaagt aattaataag gtaatTTTgg aggatttagtt	2220
ttacgtttat tacgaaatag gttaatggaa tagagaatgg taggaaggga gcgtatTTTA	2280
ttagattggg ttTTTTGTC ggattGGGTgt tggagtcgag gtttgaaggt aagaaggttt	2340
gtatTTGAA gaaagggttg aggtttcga gatTTAAGGA agtaatagag acgttagagtt	2400
tatttgcgtt atTTTATATG ttattGGTta ttgttgagtt gaaatgttagt tagTTTaaa	2460
ttaatttGtGtGtGta aaatataatat tagatTTTaa g	2501

<210> SEQ ID NO 50
<211> LENGTH: 4501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 50	
tttatttGtt ttataggatt ttatGGAA ttttggagtt ttggaggcga gagggatttt	60
ggatattatt gagTTTTatt ttatTTAA taaatataga agtggacgtt tggataggt	120
aagtgttttG attaaggtag gtgtatAGTT atTTTGtaat attggaaata aattttAGGT	180
tttttGATT ttgtttttta ttatTTTTt ttTTTatTTt ttGAAataa agTTTTatG	240
tgtttttttt tatagtgata tttttGGAat gtattAGTT gtaatTTAGG aaggaaaaaa	300
aataaaatata taagagataa atttGTTAGG aggataaatt ttttGTTTTt ttGATTTGTT	360
tagagggtGA ttattttat ggttagagaat tatttaatTA gtGtaatGAA aattttttG	420
tgggttGGGT attgtataaa gatTTAAACG aatttGTTA tagattGAA aagttagatac	480
gagatttGtG aatGGTTGGG gtttttaAGT ttatAGTATA agtATGGGTt atatTTATA	540
gtttGGAGGA ttGAGTTTG AAAATGGGTa agttttttta ttTTTTGAA ttTTTatTTT	600
tttatTTTA aaataaggat gagtagTTTt tgaggTTTTt ttacgtttt ttTTTTat	660
agattttAGT attttataat ttGATATAAA gagGGGTGGt atGAATTtT ttttttGAGA	720
aaagtTTTAG gaaAGAGAAT attAGGTtT tttagtaggt gtGtagatAG gttAGatAGA	780
ttttGAAatt tatttagttt ttTTTatGtG tataattttta ttattGTTT tagttGttAA	840
gagaaAGTAG gagAGTTGTt atttttattt ttTTTTTTt ttTTTTTTt ttGAGACG	900

-continued

-continued

tggatgttta cgggggtaag ggtttggtgt tgaaaaattt agtattttga atcggtttt 3240
ttgaatgaat atagaatttc gtttgggggg ggagagtata gaaaatagtt ttttattata 3300
tattatagtt agttgtaaat agtagatggt tttttatatt ttagagagta agaatttagag 3360
agagagagaa agagagagag tttgggggtt tttttttgt gtttgggggg ttttagagaaa 3420
ttggagggggt agtagtttgtt atttttcgt tggttttatt aagtataagtt aaggtttta 3480
ggatatgggtt attttttattt tggtggaaacg gttttgttgg ggtggggggg tgtagttgg 3540
ttttgggttg ggtagagat atttagtggt ttaggtgggc gtgggggttag ggcgttagacg 3600
agaaggggta cgagggtttc gtttcgagga tttagcggta agtatccggtt tcggggcgegt 3660
tttagtttat ttattcgcgt gtttacggcg gtattttttt ttataaggat ttgaacgatt 3720
cgggggcggt ttctgttgcgt tattttttgtt tttcggtttc gttttttttt tggagggtcg 3780
atgaggttaat gcccgtttgtt tattgggttgg agggggggggg tttaatagttt tggagggttt 3840
gttttccgggg gtttagcggtt atattattcg gtcgttttagg tagcggcgta gagegggttag 3900
taggttaggcg cggggcggtt agacgggtttt tttttttttt tttgtttttt tagttttgtt 3960
tttttgcgtg ggagggtcggtt cgtcgagttt tgcgttagcg tgcgggttagt ttcgttgcgt 4020
tttatttcgt ttctgtcggtt attcggaggg tagcgcgtcg gaggtaagg ttgttgcgt 4080
cggttcggcg ggcgagcgag ttccgggttgtt agtagttcg tccggggcgc gtacggtaat 4140
tttggagagg cgagtagtagt ttccggtagc ggccggtagta gcccgttaatgtt tttttgggtt 4200
cggtttatc gtgttttgg gtagttggag tttgggggtt tgccgggtcg aggctgtat 4260
atgttcgtttt agtttattttt aggacgtttt ttgttaatttc gatatccgtt acgttttttg 4320
gtgtttcgtt cgagttttac gttgttagta ggattgttagc gttgttttagg gagtagggc 4380
gagttttattt tttttttttt gtttaggag aggggttagac ggggttgggg cggagttggag 4440
aaattcgtatg tttttggcg gggccgttgg tatagttttagg aggggaagat gttttgtttaga 4500
g 4501

<210> SEQ ID NO 51
<211> LENGTH: 4501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

-continued

gttggtcgtt ttgcgtcggtt gtttggccgg tcgagtgata tagcgttggg ttttcgggg	660
tttcgtttcg gggttgtggg gttcgttttt ttagattaat ggtagagtcg tattattta	720
tcggttttt aaaaagggggg cggggtcggg ggtaaggggt aacggggcgg ggtcggtttc	780
ggatcgttta gattttata ggaataatg tcgtcgtgg tacgcgagtg ggtgggttgg	840
ggcgcgttcg ggatcggtgt ttgcgttgg gtttcggag cggagtttc gtgtttttt	900
tcgtttcgct tttggttta cgtttatttg ggttatttgg tggttttgcgtt taaaatttgcgt	960
attaatttata atttattttat ttttagtagga tcgttttat aggtgagggg tggttatgtt	1020
ttagagattt tgatttgtt tggtggaaatc agcgggggaa tggtaattgtt tatttttttta	1080
gttttttgg agagagtagg tatagaggaa aaagatttaa atttttttt tttttttttt	1140
ttttgggtt ttattttttgcgtt ggatatggaa agttatttgcgtt tggttttgcgtt tggatgtt	1200
atatgataga agattgtttt ttgtgttttt ttagagtaaa cggggtttttgcgtt tattttttta	1260
ataaatacgg ttttaggatgt tgggttaagt aatataaaatgtt ttttgcgttgcgtt gtggatattt	1320
atgaatgtat aggggtgattt tttgtgtatgt taggggtatgcgtt aagaaaataa aatagatgtt	1380
tgggttggat atgattgggg aggggttgcgtt ttttttgcgtt tagtaaggaa agggttttttt	1440
gaagtggtt tttcggttgcgtt aagtggataa agataaggat tagtttggg gtaagttataa	1500
ggagagaggtt attaggttgcgtt gggaaatgtt aagtgttttttgcgtt ttttttttttgcgtt gtttggatgtt	1560
tagaaagtagt aaaggtaaga ggaagtgggtt ttttgcgttgcgtt ggttgcgtt gtttgcgtt	1620
ttagtaagggg atttgggggtt gagttgttgcgtt aaaatttgcgtt taaaatttgcgtt agaaccgtt	1680
aggatttttttgcgtt acgattttttgcgtt ataatagatgtt attgttagata attaataataa ttatttttttgcgtt	1740
tatttttttgcgtt attttttttgcgtt tgtttttttgcgtt ttttttttttgcgtt aggattttttgcgtt attttttttgcgtt	1800
attatagtttgcgtt ttataataat ttttgcgttgcgtt aggaaaatgtt agtataatgtt ttattttttgcgtt	1860
tttttttttgcgtt aggaatcggtt atagaaatgtt gggatgttgcgtt tggttttttgcgtt aggaatcggtt	1920
tatttttttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt agagagatgtt gtatttttttgcgtt gaacgttgcgtt	1980
agtttttttgcgtt ttttgcgttgcgtt aatgggttgcgtt tgtagtttttgcgtt attttttttgcgtt ttttgcgttgcgtt	2040
tttttttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt aatgttataa gtttggaaataa	2100
aatagaaataa gggagaaatgtt taagtttttgcgtt ttttgcgttgcgtt gaatttttttgcgtt gtttgcgttgcgtt	2160
tagataatgtt atattgggttgcgtt agtagaggttgcgtt gaagaatttgcgtt tgagagatgtt ttttgcgttgcgtt	2220
aatattttttgcgtt aaattttttgcgtt ttttgcgttgcgtt agtatttttgcgtt ttatagaaatgtt gggaaatgtt	2280
tttttttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt	2340
tttaatttttgcgtt gtttgcgttgcgtt tagagatgtt ttttgcgttgcgtt gatgttttttgcgtt tagtatttttgcgtt	2400
tttatttttgcgtt ttttgcgttgcgtt gatgttttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt	2460
gtttgggttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt gtttgcgttgcgtt aattttttgcgtt gtttgcgttgcgtt	2520
gattggagatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt	2580
ttatatttttgcgtt ttttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt	2640
ttttcggttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt	2700
attttttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt	2760
atatttttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt	2820
taagtttttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt ttttgcgttgcgtt	2880

-continued

ttgaagcgtt	tttttttagt	agttatttaa	ttatgaaagt	ataagtataa	tttgattttt	2940
ttttagttg	aaaagtaacg	aaattatttt	ttggtaatt	ttttttattt	ttttatTTTA	3000
agttggTTTA	tatagtgttt	tgaagtttag	cgaattttaa	gagttttta	gttggatgt	3060
tttttttat	tttaaatttg	agtagtttt	ttttgttta	tcgtgtgaag	gtagttggT	3120
tttattttta	aggaaaagaa	atagtaaatt	ttatgaaattt	ttgcgttaggg	gagtttttt	3180
gttagggTTA	tttttttag	taggtattta	ttagttggA	tgttatggT	tttatgagtt	3240
taataatATG	taagaattgg	ttatTTAATA	ttatTTAATA	agttAGGTGG	GGTGAATTG	3300
aggTTAATAG	taagaatgaa	gatgttagtt	gggtgcggT	gtttacgTTT	gtatTTTAA	3360
tatTTGAGA	ggTTAAGGTG	ggcggattac	gaggTTAGGA	gatcgagatt	atTTTGGTTA	3420
atATGGTGA	atTCGTTTG	tattaataat	ataaaaaatt	agttAGGTGT	GGTGGTGGC	3480
gttttagtt	ttagTTTTT	gggaggTTGA	ggtaggagaa	tggcgtgaat	ttgggaggc	3540
gagTTTGTAG	ttagTTGAGA	ttatTTTATT	gtatTTTAA	ttaggtgtG	gagtggatt	3600
tcgttttAA	aaaaaaaaaa	aaaaaaaaaa	agaatgaaGA	tgttagTTT	tttggTTTT	3660
tttgatAGT	aagaataatG	atagAGTT	atatTTGGGA	agaattggT	aatTTTAA	3720
atTTATTTGG	tttgTTTATA	tatTTATTAG	gatgatttgg	tatTTTTT	tttggaaATT	3780
ttttagggAA	ggtgagTTA	tatTTTTT	tttggTTATA	agttatAGGA	tgttagAGTT	3840
tgttagGAAGA	gaagTCGTA	aaaggatttt	agaaattatt	tatTTTTGTT	ttaaatgtgg	3900
aaaaaaATAAG	gttttagAGAA	gtgaaggAAT	ttgttttattt	ttagggTTTA	gttttttaAG	3960
ttgttaAGGTG	tggTTTATGT	ttgtattgtG	ggTTTGGAAA	ttttagTTAT	ttatAGATT	4020
cgtatTTGTT	ttttagATT	gtagatAGAT	tcgtttGAGT	ttttgtatAG	tgttagTTT	4080
atAGAGAAAT	tttatttATA	ttgatTAAT	aatTTTTAT	tatgataATA	attatTTTT	4140
gagTTAATTA	gaagtaatat	aggtttgtt	ttttgtatGG	tttattttt	gtgtgtttat	4200
ttttttttt	ttttaaatttA	ttagTTAATG	tatTTAAAT	atattattat	aaaaaaaAGT	4260
atATGAAAT	tttgTTTTG	ggaaatgAA	agagAGTAA	gtggAAataa	aaaattaaaa	4320
gatttGAGAT	ttgtttttaA	tgttgtagAA	tagttgtGA	tttggTTTGG	ttaagtatt	4380
ttgtttGTTT	aggcgTTTAT	ttttgtgttt	atggatGAA	agatagaatt	tagtggtatt	4440
taggattttt	ttcgTTTAA	aaattttaAG	atTTTATGGG	gaattttgtA	ggataagtGA	4500
a						4501

```

<210> SEQ ID NO 52
<211> LENGTH: 3001
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

```

<400> SEQUENCE: 52

gaagtgttaa ttttagattt ttatTTTTA tataagTTA tttttgtatt agggtagtga 60
ttttttttt tgggtgagat tttgaaattt gggattataa ttttgaatta taattataaa 120
atggtatttG gttgtaaattt atTTTTTTT tttttttgtt tttttagattt gatattatgg 180
atTTTTATAA ggattttatgt tttttattta tttaatgaa tagttgtgg gtaataattt 240
tagaagagtt ttaattttta ttaggagaat ggataaggTg gagaagtaga gaaaatgtaa 300

```

-continued

tgagtagaaat gtttaagtta ttatTTTgga attgattgaa tataaataaa aatgagaaag	360
atacgtaaaa aagaaggaa tggtaagta gggatgtt tgggagagga ggggtttat	420
agttatgaga gttaatTTTg taatatTTta tagggTTata atattgttt ttatATATTg	480
aggttagtagt agggaaattt ttaattttt agaaatattt aattttttt ttatTTTTa	540
aatTTTTTT ttatTTTTTT ttatTTTTg taatTTTTt tgTTTTaaa	600
atgatTTTTT ttTTTCGGA agaagtaatt tttaaattt agtttatata agggatttg	660
atatgtttaa taagtttaa atatattgtt ttttagtaata tttagtatTTTt gtttatttg	720
agTTTgagt aatttgtatt ttaagTTtag tttttattgt tttgttttg gtAAattttt	780
attaagtgtt ttTTTTTTT aaatatacgt atatgtttt tagattttaa agTTTTTTat	840
gaatatgtaa atTTTTTTTT ttgaaaattt ttgcgtgag tggTTtagtag gttaatttt	900
ttattgttaat gtggTTTGT gttagggtt tggTTcgTg ttgTTTgtaa gataattata	960
gatgtgattt ttttttagaa gttttgaat ttttaagat agtttggTTt ataagaaaat	1020
tAAAAGGTGG aggtcggcgc cggTggTTA cgtttgtaat tttagtattt tgggaggtcg	1080
aggccggccgg attattttag gttggggagt cgaaatttagt ttgattaata tggggaaatt	1140
tcgttttGT taaaatataa aaatttagtta ggcgtggTgg tggatgtttt taattttgt	1200
tattcgggag gttgaggtag gagaatcggt tgaattcggg aggtagaggt tgcgtatgt	1260
cggatcgtg ttattgtatt ttagTTTggg taataagagc gaaattttgt tatataata	1320
taaataatata tatataatata tatacgggtt agtttaggaa gtaaaaaaaaaaaaaaaa	1380
aaaatttagat tttttttat attttagatt tgaaggataa aatttttaggg ttagggcgtt	1440
cgtttatTTta attttatatg tatttttagg ttatTTtagta tttaggtatt tagtatttag	1500
gtatattgtg gttttttatt tttacgata gtagtaataa cgttggatgg aagtTTTTta	1560
ttgtgtgtta cgggTTatgg gttatgtgt ttagaatttt atgtgaaattt aatTTTTat	1620
ttttacggat atTTTTgaaa tagatgtttt agttttttt ttgttaacga ggttagttgag	1680
gttttttagag gtttaatattt agtattatgtt gtcgtatgtac gtaaggtaaa tatacggtt	1740
ggtgagtata tatttggTTt gtattttatg cgtttaaTTta taaggTTTT ttTTTTtagg	1800
aaggTCgttG tttttttgg gacgattttg tagTTTgag gtatgtatgt acggTTTTT	1860
agaagggtga ttaggaggTTt ttttgcTTt tagTTTcgG cgtcgatTTt tattgttaggg	1920
tccgggttGt gattcgtggg gacggTTTTt tgcgtttcgG cggggggagt gggcggggag	1980
ggggcggcggc gcgtcggggc ggggTTcggg acggTcggtt tgggagttgg agtttatAGC	2040
gggaaggcggt cgtcgTTegg gttcgttagg gttaggegag gcgaggGGGGG gcccgggtcg	2100
gcgttacggg aaggggaggt cgcgcggatc gggagtcgta tcgcgttagt cgggtttag	2160
cggTCgtcgtta ttaagggtgc gatggggTTt gagacggaga aggccggacgt atagTTTTT	2220
atggacgacg atTTTTatAGC ttattatAGC ggttTcgagt acgtcgatTTt cgagaagttc	2280
gcggattcgg attaggatcg ggattttat cggTTtaatt cgtattttaa ggtgaagttc	2340
ggggcggggcg ggTTtaagtt ttcgtttagg tccggaggtg cgggcgtttt ttagTTTcgt	2400
ttaattcgt ttattttatg ttatcgggtc ggttTcgtag ggtttgagat tcgtattttt	2460
tttcggTTTT attcgTTatt aggtcgtcg cgttagttagg aatttttagt taggtttttt	2520
tgcgtttatc gtgattttaa gagaagaggc ggacgttttTt gtaCGTTTT ttTTTTgtt	2580

-continued

tttttcgttt aaagcgaaaa cggtttcgg ggcgttaggt tggttgatag ttcggggttt	2640
ttgcgttttg ttttttagt tgggtttcga ggatgtgatc gtagagtccgg tgattacgta	2700
ttttttgat aaagtgtgaa tttgttagtt aaaaaaaaaaaatacgtaat	2760
gtataagttt ttgacggtgt ttttggttat tttttgggtt ttttattgcgg gaattttttt	2820
tgttattttt agttgtttgt atatttggtg agacggggta tatcgggtgg atcggttttt	2880
tgaaaatatgg gtatattttt cgtttattgt tttttatttt ttttttattt tagtccggcgt	2940
ttagaggaggaa acgcgtat tagttttaa gtagtaggaa gaatttggaaag gttttgaaag	3000
q	3001

```
<210> SEQ ID NO 53
<211> LENGTH: 3001
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
```

<400> SEQUENCE: E3

-continued

tttgagtgtt aggtatttga gtgttaggtg atttgtaagt gtatgtggag ttgggtaggc	1560
gaacgaaaaa gttttaggtt ttgtgtttt agatttgggg tgtgagggga gatttgattt	1620
tttttttttt tttttttta ttttttaaat tatatacggt gtgtgtgtgt gtgtgtgtt	1680
gtgtgtgtgt ggttaggtt cgttttgtt gtttaggtt gagtgtaaatg gtacgatttc	1740
gttttatcgat aattttgtt tttcggtttt aagcgatttt tttgttttag ttttcgagt	1800
agttgggatt ataggtatgt attattacgt ttggtaat ttgttatttt agtagagacg	1860
gggtttttttt atgttggta gggtggtttca gatttttaa ttttaggtga ttcttcgtt	1920
tcggttttttt aaagtgttgg gattgttaggc gtgagtttgc gcgttcgatt tttttttttt	1980
aatttttttg tgaatttagat tgtttgaaa gatttaggaa ttttaagat gtagttat	2040
ttgtgattat tttgttaggtt gtacggaaat agaattttaa tataaaagttt tattgtat	2100
gatgaattttt tttgtttagt atttacgtt aagtttttag aaggaggag tttgtatgtt	2160
tataaagggt ttttaggtttt ggttagatata tacgtgtgtt tggggaggaa gggttatttt	2220
tgaaaatttg taaaagtaa aataatgaga attagggtt aggtgttagt tgtttaggt	2280
ttaaaagttagg tatgttagtag gtattgttgg atatagtgtt tttggagttt gttgaatata	2340
ttaaattttt ttatgtgaat tggatttgaa gaattgtttt tttcgaaaga aaaaagatta	2400
ttttgaaat agtaaaaattt atagaatataa aaaagaatag gaattgaatg agaaaaatgt	2460
ttgggggtgg gaggttaagt ttaatatttt taataattaa aaagttttt ttttgttatt	2520
ttagtataatg aagggttagt ttgtatattt atagggttgg atagagtttga tttttatgtt	2580
tatggagttt tttttttttt agatattttt ttgttttattt atttttttt tttttacgtt	2640
ttttttttat ttttattttt gtttagttaa ttttaaagtgtt atgattttaga tatttttattt	2700
attgtatattt ttttgttttt ttatgtgtt tattttttt atgagaattt gaaattttttt	2760
agaattgttg ttaataattt gtttattttttaa gtgaatagaa gatatgaatt tttatggaa	2820
tttataatataat taattgttagt gaaatagaaaa aaaaaggaga taattttagt ttaaatat	2880
tttataattt ataattttaa attataattt tagattttaa gttttttttt aaagaagaaaa	2940
gttattgttt tagtataaga gtgggtttt gtagtgggtt aaaaattttagt attagtattt	3000
t	3001

```

<210> SEQ ID NO 54
<211> LENGTH: 2501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

```

<400> SEQUENCE: 54

tttggatttt aatgtgtatt ttatattttt agtataattt atttgggatt agttatattt
tagtttaata atagtttaata gtatatggga tagtgttaat aaattttgtt tttttgtt
ttttttgggt tttggagatt ttaatttttt ttttagattt taaatttttt tttttttttt
ttttgggtttt aatatttagtt tggtagagga atttagtttta atgaggtatg tttttttttt
gttattttttt attttattaa tttgtttttt ggttaatgtt ggtttttttt tttttttttt
tttttattttat tagtttataat atttattttt tattttgtt attagaatgt aggtttttgg
aaggtagggaa tttaaaaaaaa tttgtttttt tttatgtt gttttttttt tttttttttt

```

-continued

gtttggata agttgggatt ttagtatata ttttggatg gaagaattgt gtttttttag	480
aatttagtta gagggttagtt tagtaatgtg ttataggtgg ggtgtttgtg ttttgggtgg	540
atgtatttgtt ttttgggttg gtgtgggtgt ggggtgagtg ggtgtgtgtg ggggtgtgt	600
ggtagagtgt gttagtgagt ttggagtgtg gagttgggag gagtagtgag tgttgttag	660
aatttgtatgtt gttgggttgg tagggtagtt tggaggtggg tgggttgtgt tgtagttt	720
ttttaggggt ttttatttgtt tggttgttgg ttgtttttgt tttaaaaggt ggtaaggagt	780
tgagagggtt ttttggagtg tgaggaggat agttggattt agttaatgtt ggggatttt	840
tttttttgtt ggaggggatt tggttaatttg tagtggtagg gtttgggtt ggtgtttggg	900
agggattttgtt gttttttattt tatttttagt tggtgttttt gttgttgtt tggttagttt	960
ttaggttgttgg tttttatgtt tggtttttga tagtgttttgg gagggatttgg gggagttttt	1020
agggtgtttgg gtgagtagtt aggtgtggtt ttttggtttt tttgattttt ggtgttagtt	1080
ttaggtttttt tagtaggggt gtgtgggggt ttgtttgggtt agtgttttga gtaattggga	1140
aggtaagggtt ggagggaaat ttgggtttgg ggagaagtgat gattgttagtt gggagtttt	1200
ttaggttttttgg tgggtgggtt gagaatagggtt ggtgtttgggt tgatttaggtt ttttgggtt	1260
gggtgtgagggtt atttggagtg aattgtgttggg ttttgggtggg ttgtttttt ttttttttt	1320
ttttttgggg tgggtgtatgtt ttgggttgggt tggttaatgg agaggggattt gtttaggaatg	1380
tgggtttgggg gattgttttg tttggggaaag gggaggggggtt ggttatgggt tgtaggagg	1440
tgtggggagttt gagaggggtgtt gtgggggtgtt tattgtgtt tgtaggttggg gagagattgt	1500
ttaggtgtatgttggg tttgggttggg ggttttatttt ttttgggtgtt gttttggagg	1560
tggaaagttt ggggtgggtat gtgtgttagat aaagggggagg tggtgggtt agttttttt	1620
ggaggtgggtt ttgattttaa tagattgtttt agtgtttggg gatgttgatt ttgggtgtt	1680
gtgggttttgg tttttatgtt tggtgtgggtt tgaggggtgtt ggggtgtttt tggtgttttta	1740
gttttaataaa aggggtttttt tttttatgtt tgagggagggtt gtagttttgg agatttggtt	1800
ttaggtgtatgttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	1860
tattattttt gttttttttt tttttttaaa aggattggag attgatgtat gagggggtta	1920
tggaggtgtatgttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	1980
gtgaggtgtatgttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2040
atattgggtatgttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2100
gaaatgtata atgatggggatgtt ttttttttttggg ttttttttttggg gtttattttat	2160
tttttttttttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2220
tgttttgggtt ttttttttttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2280
gtatagttata ttttttttttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2340
tgttttgggtt ttttttttttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2400
gttagttttt ttttttttttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2460
ttagtattttt ttttttttttggg ttttttttttggg gtttattttt ttttgggtt gtttattttat	2501

<210> SEQ ID NO 55

<211> LENGTH: 2501

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 55

ttaaattata ttgaaaattt agatagaggg taaagtatta agattttaa aatttattaa	60
ttagttgtta tttttgttt aaattatata taaaagttaa ttttaattgt aattataatt	120
ttagtgggg ttttgtgtt atagaagggt aagggttgat aagagttgt atggataatt	180
aaatttgatt ttgttaggg gattggatta atatattgtg tattataatt tagtttgaa	240
ttgttgttt ttttttattt ttaggtttt ggatttaggtg gtttattgtt gggtagtaa	300
ttttttttt tatgatttgat gtatttttag aaagggtgag ggagatgtt gatttaattt	360
gtatttgttt attatataat ggtttattt ttagtattt ttataattt ggagggtatg	420
ttttggat ttttttattt tgtaagttt ttgtgtgggt tgtagtaat tatttatgtt	480
ttttgtttt aatatggta tgataaaattt ttatgtttt taaaagttaa gggattttt	540
ttttgtttt ttagattttt attattgttt ttgtgtttt gtatgtttt tatgtttag	600
tttttagttt ttttgaagaa aaaaaaaagg taataataat agtaaaatgg taattttaaa	660
aggaagttaga tttttttgtt gttaagattt tggtgttga agattgggtt ttggagttt	720
ttttttttt tgggggtggag aggagtattt ttgttaaag ttgggtgggtt agggatgtt	780
ttgattttttt agttttgtgt gtgtgtgggg ttgggtattt tggttattttt aggttgggt	840
ttttaatgt tgggtgggtt gttgggatta gaattgggtt tggaagtgtat tggtttttt	900
gtttttttt ttttgtata tatgtttattt tagttttttt attttttaggg atgttggaa	960
agaatgaagt ttttgtttt aggtatgtg ttttattttt agtaattttt ttttagttt	1020
tggtaatgggt ggtattttttt tggtttttt tggtttttt gtttttttta atattgtgtt	1080
tatttttttt ttttttttga gtgaggtagt ttttagagtt atatttttgg tgattttttt	1140
tttgttattt gggtgattttt gtgtgtgggtt gttttggggta gtaaggaggaa gaaggaggatg	1200
gtttattttagt gtgttagtagt ttgttttaga tttttgtgtt ttgtatataaa gtgtttgggt	1260
gggttgggtgt tatttgggtt tatttgggtt tggttttttgg ggaagttttt tggtgtat	1320
tgtattttttt tttgaagttt agttttttt tgttttgggtt ttttagtttgg tttgaggtat	1380
tgtttggata aattttattt tgttttgggtt gggaaagttt aagttgggtt tggttttttgg	1440
ggggattttttt gagttgtgtt tggttattt tttgggtgtt tggttttttgg tttgggtttt	1500
ttggaggtttt gttggaggtt gattatagggtt gttagtgtgtt gagatttttttgg ggggttttgg	1560
tgggaatata gtttagggagt gagtgggggg tgtagattttt ttttaggtgtt tggttttaga	1620
ttttgtgtttt tgtaggtttt gatgtttttt tggtggaggaa ataaagtttt tggtgtgtt	1680
ttgggtttgggt tgttttttttt atattttgaa gtagttttttt ggtttttttt tgttttttgg	1740
gtggagggtt gttggtaggtt ggttaatgggg gattttgtaa tggttttttgg ggtgttttttgg	1800
atttttttt gatgtttttt gtttaggtttt tggtgtgggtt tttgtgtgtt gttgtgttttgg	1860
tttttttagttt tggtgtttttt tggtgtttttt tggtgtatattt ttgtatataat	1920
ttatattgtttt tattttttatg ttgggtgggg agtttagtgtt tttgttttttgg atgtgggtgtt	1980
tttattttgtt atatattgtt aagttgtttt ttgtgtgggtt ttttagggaaa tatgttttttgg	2040
ttgttttttagt atgtgtattt ggttttttagt ttgtgtgggtt ttttagggaaa tatgttttttgg	2100
aattatataatg aataaaatag atttttttaa atttttgtttt ttggaaattt tgtagttttgg	2160

-continued

tggatagat agataataaa tatgttaagt aattaataag gtaattttgg aggatttagtt	2220
ttatgtttat tatgaaatag gtaatggaa tagagaatgg taggaaggga gtgtatttt	2280
ttagattggg ttttttgtt ggattgggtgt tggagtttag gtttgaaggt aagaaggttt	2340
gtaatttgaa gaaagggttg aggttttga gatttaaaga agtaatagag atgttagagtt	2400
tatttgtgtt attttatatg ttattggta ttgttgagtt gaaatgttagt tagtttaaa	2460
ttaatttgtgt tgtaagtgtta aaatatataat tagagtttaa g	2501

<210> SEQ ID NO 56
<211> LENGTH: 4501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 56	
ttttttgtt ttataggatt ttttatggaa ttttggagtt tttgagggtga gagggatttt	60
gcatatttttattt ttttatttaa taaatataga agtggatgtt tggataggtta	120
aagtgttttgc attaaggtag gtgtatgtt attttgcattt atttggaaata aatttttagt	180
ttttgttattt tttgttttta ttttattttt tttttatttt ttagaaataa agttttatgt	240
tgtttttttt tatagtgata tggggaaat gtatttagtta gtaattttagg aaggaaaaaa	300
aataaatata taagagataa atttgttagg aggataaattt ttttgcattt ttgttgcattt	360
tagagggtga ttattttat tttttttttt tttttttttt ttagaaataa agttttatgt	420
tgggttgggtt attgtataaa gatttaaatg aatttgcattt tagatttgc aagttagat	480
gagattttgttgc aatgggtttttt gttttttttt ttatgttata agtattttttt atattttata	540
gtttggagga ttgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt	600
tttatatttttta aaataaggat gagtagttttt tgagggtttt tttatgttattt tttttttttt	660
agatttttagt attttataat ttgtatataaa gaggggtttt atgtttttt tttttttttt	720
aaagtttttagt gaaagagaat attagggtttt tttttttttt tttttttttt tttttttttt	780
ttttgaaattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	840
gagaaaatgttgc gagttttttt tttttttttt tttttttttt tttttttttt tttttttttt	900
gagttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	960
ttgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1020
tgtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1080
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1140
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1200
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1260
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1320
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1380
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1440
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1500
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1560
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1620

-continued

gatgtttttt atgagaaaat tgggtttga agtttagtag aagttattta tttgatattt	1680
atattttagg taagggtttt tggtggagaa aatattggta ttggataaa aattgaaatg	1740
tgaaaagaaa gggaaagagag gggttttatt atgtaagatg ttatattaa gtggatttg	1800
tttggaaagt tttttaaat tttttatatg attgtggaat aagttatgtg gggtgtgggg	1860
ataagtgaat tttttaatt ttattatgtt tggttttatt taatttgat ttttagagtg	1920
gttttaggg tttttgttt aggatttatg tagttgttg ttatattat gtttttagt	1980
tttttgagat ttatattggt tttgagaggg taaaaagta gtgtggtaa atattttagg	2040
ttttaaagta ttttattgtt gggtggggaa gtaatagaat tatattttat aaaataatga	2100
aaatagtgtt agaaaaatat tgagagatag aaatattttt atgagtttagg ttatagttag	2160
agtgaaggta gggaaaggttt taaaagttgg gtggagggga taagttaaaa agatgtggaa	2220
attgggtttt tttttttagt gttaaagtgt taaaaggggaa aaaaggagtt taaaatgt	2280
ttttggaaat attatttttt atgaattttt tggttttgt tggtaatg ttattttgtt	2340
gagatgtaaa tagaggagtt ttgagaaaaga agttgaattt gtatttttt ttgttttat	2400
ttgttttaaa ttgtgttat tttaatagg atgaagtggaa agagaaaaggg aaagagataa	2460
aagtgttagaa agatggaaga ttttagttgt aaatggttat ttgttagttag atggaatagt	2520
tgttgatgtt tagggaaatg tatgtttttt tttagatggg aaggagtagt ggaaagggg	2580
gatgatgtt tgggtggta ttaatttattt tattttttt tggtggttt ttattttggaa	2640
agtggaggtg atatttgtgt ttgtttttt tattttataaa gattttgtg agagttataa	2700
tatggtgaga tatagaattt tgttttaaa aatataaagt agaattaaga tggtaataat	2760
aaggatagta attgtgttag ttatttgtaa ttattttat tagtttagtg tttaggattt	2820
tggattgtt ttgtgtttt attatagttt tggatttagt tattttaaa ttgtgtttt	2880
aagggtggag ttgtgttagt tatggtagg gaatttttt ttgtgtttt ttatttttt	2940
gtttttaaa tatgttttagg gttttgtat ttgtgtttt ttgtgtttt tattttttt	3000
ttgtgtttt tttagagtt gatttttgtt ttgtttatt tttagtgag gatgttattt	3060
tagggagttt ttttttattt attgttagaga gagtaggtt tttaggttta tggttaattt	3120
agaattttgt ttgtttttt ttatgtttt agtattatag aaaatttattt tggttattta	3180
tggatgtta tggggtaag ggtttgtgt tggttaattt agtattttga attgtgttt	3240
ttgaatgaat atagaatttt gtttgtttt ggagagatata gaaaatagtt ttattttata	3300
tattatagtt agttgttaat agtagatgtt tttagatattt ttagagatg agaatttagag	3360
agagagagaa agagagagag ttgggtttt ttgttttgtt gtttgggtttt tttagagaaa	3420
ttggaggggt agtagtttagt attttttgt tgggtttatt aagtagttt aagggtttta	3480
ggatatggtt attttttattt tggtaatgtt gtttgggttgg ggtgggtggg tggttagtt	3540
ttttgggttgg ggttagagat atttagtggt ttaggtgggt tggtgggttag ggtgttagat	3600
agaagggtta tgagggtttt gtttggagaa tttagtggta agtattgggt ttgggtgtgt	3660
tttagtttat ttatttgtgt gttttaggtt gttttaggtt ttataaggat ttgaatgatt	3720
ttgggggtgt ttgtgtttgt tattttttt ttgtgtttt gtttgggttggg tggagggttg	3780
atgaggtaat gtgggtttgt tattgggttgg aggggggtggg ttttaatagt ttgaggtggg	3840
gtttttgggg gtttagtgtt atattatttg gttgtttagg tagtgggtgtt gagggttag	3900

-continued

taggttaggtg	gtgggtgttt	agatggttt	ttttttttt	tttggttttt	tagttttgt	3960
ttttttgtt	ggagggtgtt	tgtttagtt	tgtgttagt	ttgaggtgt	tttgttgt	4020
tttattttgt	tttgggtgggt	atttggaggg	tagtgtgt	gaggtaagg	ttgtttgt	4080
tgggttgggt	ggtgagtgag	tttgggtgt	agtagttt	ttgggtgt	gtatggtaat	4140
tttggagagg	ttagttagtag	tttggtagt	ggtggtagt	gtggtaatga	ttttttgtt	4200
tgggttatt	gtgttttgg	gtagttggag	tttggggat	tgggtgtt	agggtgttat	4260
atgtttgtt	agttatttt	aggatgttt	ttgtatattt	gatattggta	agtgtttt	4320
gtgtttgtt	ttagtttat	gtttagtta	ggattttagt	gttgtttag	gaggtaggt	4380
gagttttatt	ttttttttt	gttttaggag	aggggttagat	gggggtgggg	tggagtggag	4440
aaatttgatg	tttttgggt	gggggtgtt	tatagtttag	aggggaagat	gttttgatg	4500
g						4501

<210> SEQ ID NO 57
<211> LENGTH: 4501
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 57						
ttttgttaggg	tatTTTTTT	ttttagttat	gttagtgttt	ttgtttaagg	atattgagtt	60
tttttatTTT	gttttaattt	tgtttgttt	ttttttgggg	tagagggaaag	gagtgggggtt	120
tgtttgttt	ttttaaatTT	tttttttttt	tttttttttt	tgtgggtttt	gggtggggta	180
ttaggagtgt	ttattgtatgt	tggagttgt	gaagggtgtt	tgggggtgtt	tgggtgagta	240
tgtgtatgtt	ttgggtgttt	agtttttag	gttttagtt	tttaggagta	tgtatgttt	300
gagtttaaggg	gttattgtt	ttgttgtt	ttgttgtt	ttgttgtt	gtttttttaa	360
agttgttgt	tgtgttgtt	gtgggggtt	tgtatgtt	ttgttgtt	ttgttgggtt	420
gtgtgggtta	atTTTGGTT	ttgtgtgtt	gtttttttag	tgtttgtt	gtatggatgg	480
ggtgtatgt	gtttgtttt	gtgtgtgt	aggatttgg	gggtgggtt	ttgggtgaagg	540
agttaggagtt	ggaggaggtaa	gaggaggagg	agaagtgtt	ttagtgttt	ttgtttgtt	600
gttggTTT	ttgtgtgtt	gtttgggtgg	ttgagtgtata	tagtgttgg	tttttggga	660
ttttgtttt	gggtgttgg	gtttgtttt	ttagattaat	ggtagagtt	tattatTTA	720
ttggTTTT	aaaaagggggg	tgggggtgg	ggttaagggt	aatgggtgg	ggttgtttt	780
ggattgttta	atTTTTATA	gggaataatg	ttgttgtgg	tatgtgagtg	ggtgggttgg	840
ggtgtgtttt	ggattgggtgt	ttgttgttgg	gtttttggag	tggagttttt	gtgtttttt	900
ttgtttgtgt	ttgggtttta	tgtttattt	ggttatttgg	tgtttttgt	ttaaaatttta	960
atTAATTAA	atttatttat	tttagtagga	ttgtttttat	aggtgagggg	tggttatgtt	1020
ttagagattt	ttagtgtgtt	tggtggaaat	agtgggggaa	tgttaatgt	tatTTTTTA	1080
tttttttgg	agagagtagg	tatagaggag	aaagattaa	atTTTTTT	ttttttttt	1140
ttttgggtt	ttatTTTTG	ggatatggaa	agttatttgt	tgttttagt	tggttatgt	1200
atATGATAGA	agattgtttt	ttgtgtttt	ttagactaaa	tggggttt	tatTTTTTA	1260
ataaaatatgg	tttaggatgt	tgggttaagt	aatataaaat	ttttgtttt	gtggatattt	1320

-continued

atgaatgtat	agggtgattt	tttgtgatgt	tagggttatg	aagaaaataa	aatagagttt	1380
tgggttggat	atgattgggg	agggtttgtt	ttttttgtga	tagtaaggga	agggtttttt	1440
gaagtgttat	ttttgttag	aagtggataa	agataaggat	tagtttggg	gtaagtata	1500
ggagagaggt	attaggtagg	ggaatagta	agtgtaaaga	ttttgggtat	gtttgaaaga	1560
tagaaagtag	aaaggtaaga	ggaagtgggt	ttttgtttat	ggttgataga	gttttatttt	1620
ttagtaaggg	atttgggggt	gagttgattt	aaaattgtag	taaaatttag	agaatgattt	1680
aggattttag	atgatttagtt	ataatagatg	attgtagata	attaatataa	tttatttttt	1740
tattattgtat	attttgcattt	tgtttgtat	ttttaaaagt	aggattttgt	attttattgt	1800
attatagttt	ttataataat	ttttgtgggt	aggaaaagta	agtataagta	ttatTTTtat	1860
tttttagatg	aggaatttgg	atagaaagat	gggatgattt	gtgggttagtt	aggaatttgg	1920
tatttttttt	tattgttttt	ttttatttga	agagagat	gtattttttt	gaatgttagt	1980
agttgtttta	ttaattgtat	aatggttatt	tgtagttggg	atTTTTatt	tttttatatt	2040
tttgggggg	tttttttttt	ttttgtttta	ttttgttaga	aatgttataa	gtttgaaata	2100
aatagaaata	gggagaaaatg	taagtttagt	ttttttttta	gaattttttt	gtttatattt	2160
tagataagtg	atattgggat	agtagaggtt	gaagaattt	tgagagatgg	tatttttaag	2220
aatatttttg	aaatTTTTT	tttttttttg	agtattttag	ttataggaaa	gggaaattag	2280
tttttatatt	tttttgattt	gtttttttta	tttagttttta	aaaattttt	ttgtttttat	2340
tttaattgtg	gtttaatttg	tagagatgtt	tttgggggg	gatgtttttt	tagtattgtt	2400
tttattgttt	tatggggat	gatTTTTatt	tttttttaat	tatagtagga	atattttgag	2460
gtttgggata	tttagttata	ttttttttta	gtttttttag	aattaaatag	gttttttagga	2520
gattggagag	tatgggtgt	gttaatagt	gattgagttt	tgtagtaggt	gttttggagg	2580
ttatTTTtaag	gatTTTTatt	gaatgggggt	atatgtgggt	gaatttgaga	gattttttta	2640
tttttgcattt	ttatATGATT	tatTTTatAG	ttatgtggaa	ggTTTTtagaa	gatttttttag	2700
atTTTTatTA	tttggataa	gtattttata	tgatagaggt	ttttttttt	tttttttttt	2760
atattttatgt	tttggggaaa	gtgttgat	tttttttaat	ggaaggTTT	gtttgaaata	2820
taagtattaa	gtagataatt	tttggtaat	tttaagtagt	agttttttta	tagaaagtt	2880
ttgaagtgtt	tttttttagt	agttatttaa	ttatgaaagt	ataagtataa	tttgattttt	2940
tttttagttt	aaaagtaatg	aaattttttt	ttgggttaatt	ttttttatt	ttttatTTta	3000
agttggttta	tatagtgtt	tgaagtttag	tgaattttaa	gagttttta	gttgggatgt	3060
ttttttttat	tttaaatttg	agtttagttt	ttttttttta	ttgtgtgaag	gttagttgg	3120
ttttatTTta	aggaaaagaa	atagtaaatt	ttatgaattt	ttgtgttaggg	gagtttttt	3180
gttagggat	ttttttttag	taggtattta	ttagtttgg	tgttatgggt	tttataggtt	3240
taataatatg	taagaattgg	ttatTTtaata	ttatTTtaata	agtttaggtgg	ggtgaatttg	3300
aggTTTatAG	taagaatgaa	gtgttaggtt	gggtgtgggt	gtttatgtt	gtaattttag	3360
tatTTTgaga	ggTTAAGGTG	ggTGGattat	gaggttagga	gattgagatt	atTTTgggtt	3420
atatgggtaa	atTTTgttt	tattaataat	ataaaaaatt	agtttaggtgt	ggTGGTGGGT	3480
gtttgttagt	ttagtttattt	gggagggtga	ggtaggagaa	tgggtgtgaat	ttgggaggtg	3540
gagtttgcatt	ttaggtttagt	ttatgtttat	gtattttatgt	tgatgttagt	gagtggagatt	3600

-continued

ttgttttaa aaaaaaaaaaa aaaaaaaaaa agaatgaaga tgttaggttt tttgtttttt	3660
tttgatagtt aagaataatg atagagtat atatggaa agaattttagt aagtttaag	3720
atttatggg ttgtttata tatatttag gatgattgg tttttttt ttggaaattt	3780
ttttagaaaa ggtgagttt ttttatttt ttgttattt agttataggaa tgtaggtt	3840
tgttaggaaga gaagttgtaa aaaggatttt agaaatttatt tttttttt tttaatgtgg	3900
aaaaaaaataag gtttagagaa gtgaaggaaat ttgtttttt ttgggtttaa gtttttaag	3960
ttgttaagggtg tggtttatgt ttgtattgtg ggtttggaaa tttagttat ttatagattt	4020
tgtatgggtt tttagattt gtagatagat ttgttttagt tttttgtatag tgtagttt	4080
atagagaaat ttatattata ttgattaaat aattttttat tatgataata attattttt	4140
gagttttaatta gaagtaatat aggtttgtt tttgtatagg tttttttt gtgtgtttat	4200
ttttttttt tttaaattta ttgttaatg ttttttaat atattttat aaaaaaaaaagt	4260
atatgaaaat ttgtttttt gggaaatgaaa agagagtaaa gtggaaataaa aaaattaaaa	4320
gatttgagat ttgtttttaa tggtttagaa tagttgtgt ttttttttgg ttaagttatt	4380
ttgtttgtttt aggtgtttat ttgtttagttt attggatgaa agatagaatt tagtggatt	4440
taggattttt ttgtttttaa aaatttttaag attttatggg gaattttgtt ggataagtga	4500
a	4501

<210> SEQ_ID NO 58
<211> LENGTH: 3001
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 58	
gaagtgttaa tgtagattt ttatattata tataagttt tttttgtatt agggtagtga	60
ttttttttt tgggtgagat tttgaaattt gggattataa ttttgaatta taattataaa	120
atggtattt gttgtaaattt atttttttt ttttttttgg tttttagttt gatattatgg	180
attttataa ggatttatgt ttttattta ttttaatgaa tagttgtgg gtaataattt	240
tagaagagtt ttaattttta ttaggagaat ggataagggt gagaagtaga gaaaatgtaa	300
ttagttagaa gtttaagtta ttattttggaa attgattgaa tataaataaa aatgagaaag	360
atatgtaaaa aagaaggaa tgggtaaatg ggggtatgtt tgggagaggaa ggggtttat	420
agttatgaga gttaaattttt taatattttt tagggattata atttgtttt ttatattttt	480
aggttagtagt agggaaattt ttaattttt agaaatattt gttttttttt ttatattttt	540
aatatttttt ttatattttt ttgtttttt ttatattttt taattttttat tgttttaaaa	600
atgatttttt ttgttttttggaa agaagtaattt ttttaattttt agttttatata aggggatttt	660
atatgtttaa taagttttaa atatattgtt ttttagtataa ttatattat gttttttttt	720
agttttgtttt aattttttt tttttttttt aatattatgtt atatgtttt tagattttaa agttttttat	780
aatatgtttaa atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	840
ttatattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	900
ttatattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	960
ttatattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	1020

-continued

taaaaggtgg	aggtgtgggtg	tggtggttta	tgtttgtaat	tttagtattt	tgggaggttg	1080
aggtgtgggtg	attatttgag	gttgggagtt	tgaaattagt	ttgattaata	tggggaaatt	1140
ttgttttgt	taaaaatata	aaatttagtt	ggtgtgggtg	tgtatgttt	taatttagt	1200
tatttggag	gttgaggtag	gagaattgtt	tgaatttggg	aggttagaggt	tgtgtatgt	1260
tgagattgtg	ttattgtatt	ttagtttggg	taataaggt	gaaattttgt	tatatatata	1320
taaatatata	tatatatata	tatatgtgt	agtttagaa	gtaaaaaaaaa	aaaaaaaaaa	1380
aaaatttagat	tttttttat	attttagatt	tgaaggata	aatttttaggg	ttaggggttt	1440
tgtttatcta	attttatgt	tatgtttagg	ttattttagt	tttaggtatt	tagtatttag	1500
gtatattgtg	gtttttattt	ttttatgata	gtatgtata	tgttgattgg	aagtttatta	1560
ttgtgtgtta	tgggttatgg	gttatgtgt	tttagaatttt	atgtgaaattt	aatatthaat	1620
ttttatggat	attttgaaa	tagatgttt	agttttattt	ttgttaatga	ggttagttgag	1680
gttttttagag	gtttaatattt	agtattatgt	gtttagttagt	gtaaggtaaa	tatgttggaa	1740
ggtgagtata	tatgtttttt	gtatttatgt	tgttaatta	taaggttttt	tttttttagg	1800
aagggtgtt	ttttttttgg	gtatgtgtt	tagttttgag	gtatgtatgt	atgggttttt	1860
agaagggtga	ttaggagggtt	ttttttgttt	tagttgttgg	tgttgttgg	tattgttaggg	1920
tttgggttgt	gatttgggg	gatggttttt	tgtgtttttt	tgggggggg	gggtggggag	1980
gggtgggtgg	gtgttgggg	ggggtttggg	atgggtgggt	tgggagttgg	agtttatagt	2040
gggaagtgtt	tgtgtttttt	gtttgttagg	gttaggtgag	gtgaggggggg	gtgggggttgg	2100
gtgttatggg	aaggggaggt	tgtgtggatt	gggagttgt	ttgtgttagt	tgggtttag	2160
tgggtgtgt	ttaagggtgt	gatgggtttt	gagatggaga	aggtggatgt	atagtttttt	2220
atggatgtat	attttatag	ttattatagt	ggtttttagt	atgtttagtt	tgagaagttt	2280
gtggattttgg	attaggattt	ggatttttat	tgttttaatt	tgtattttaa	ggtgaagttt	2340
gggggtgggtg	ggtttaagtt	tttggggagg	ttggggagggt	tgggtgtttt	ttagttttgt	2400
tttaatttgt	tttattattt	ttattgggtt	ggttttgtag	ggttttagat	ttgtattttt	2460
ttttggtttt	atttggttt	aggttggttt	tgttagtttag	aatttttagt	tagtttttg	2520
tgtgtttatt	tgattttaa	gagaagggat	ggatgtttt	gtatgttttt	ttttttgtt	2580
ttttttgttt	aaagtgtttt	tggtttttgg	ggtgttagt	tggttgatag	tttggggttt	2640
ttgtgttttg	tttttttagt	tgggttttga	ggatgtgatt	gtagagttgg	tgattatgt	2700
ttttttgtat	aaagtgttta	tttgttagtta	tgtttttttt	gaaatttagta	aatatgtaat	2760
gtataagttt	ttgtatgtgt	ttttgggtat	ttttttgggt	tttattgtgg	gaattttttt	2820
tgttattttt	agttgtttgt	atatttgggt	agatggggta	tattgggtgg	attggttttt	2880
tgaaatatgg	gtatattttt	tgttattttt	tttttttttt	taggttgg	tttttttttt	2940
ttaggaggag	gaatgtgtat	tagttttaa	gtatgttagaa	gaattggaaag	gttttgaaag	3000
g						3001

```

<210> SEQ ID NO 59
<211> LENGTH: 3001
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

-continued

```
<400> SEQUENCE: 59
tttttaagg ttttttagtt ttttttattt tttgggaaattt gatgtgtgtt ttttttttg 60
atgttggttt gggataagag gagagtaggg ggttaggtggt ggagaatatg tttatgttt 120
agaaaagtgg tttatgggt gtgtttgtt ttattagatg tgttagatgt tgaggggtgg 180
aaagagaattt tttgtaatga aggttagggg aatggtagg aatatttta ggaatttgt 240
tattatgtat ttgttgcattt taaagagggt atgggttgcattt atttataattt tgtaaaagga 300
gtgtgttagttt attgggttttgcattt tgattatattt tttgaagttt agttgaggag ataggatgt 360
gggattttgcattt attgttagttt aatttgcattt tttgggaaattt gggagtgtttt tgggtggggg 420
aagtagggagg gaaggatgtt ttaggggtttt tggtgggttta tggtgggtgt 480
ataggaattt ggttaagaat tttgggttat tttgggtggtt tggtgtatggg tgggatttggg 540
gaagggtgtt ggttttagat tttgtggggt tgatttgcattt gtaatgggtt gatgggttag 600
ggtgggggtt ggggtgtttt gtatttttttgcattt gtttttagtgg ggattttgggtt tttttttttt 660
tgggttttat tttgagatgtt gtttttttttgcattt ggtggggattt ttgggttttgcattt 720
tgaattttttt ggggtgggtt ttttgcattt ttttttttttgcattt gtttttttttgcattt 780
tgaagagttt tatgttttttgcattt ttttttttttgcattt ttttttttttgcattt ttttttttttgcattt 840
gtttagttt ggttgggtgtt gtttttttttgcattt ttttttttttgcattt gtttttttttgcattt 900
tttgggttttgcattt ttttttttttgcattt ttttttttttgcattt ttttttttttgcattt 960
tgggtgggtt ttttttttttgcattt agtttttttttgcattt ttttttttttgcattt ttttttttttgcattt 1020
tttttttttgcattt ttttttttttgcattt gtttttttttgcattt agggaaatttgcattt ttttttttttgcattt ttttttttttgcattt 1080
gttttttttgcattt ggggtgggtat ttttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1140
tgggggttttgcattt ttttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1200
tttttttttgcattt ggggttttttgcattt gtttttttttgcattt gtttttttttgcattt 1260
tttttttttgcattt gtttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1320
tttttttttgcattt gtttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1380
aattaaatgtt taattttataa taaaatttttata atatatatgg tttatggttt gtaatata 1440
gtgataaaattt ttaattttataa taaaatttttata atatatatgg tttatggttt gtaatata 1500
tttgagttt ggtttaggtt gtttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1560
gaatgttttgcattt gtttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1620
tttttttttgcattt ttttttttttgcattt ttttttttttgcattt ttttttttttgcattt 1680
gtgtgtgtgtt ggttaggtttt gtttttttttgcattt gtttttttttgcattt gtttttttttgcattt 1740
gttttttttgcattt aatttttttttgcattt ttttttttttgcattt aatttttttttgcattt 1800
agttgggattt ataggatgtt attttatgtt ttttttttttgcattt ttttttttttgcattt 1860
gggttttttttgcattt ggttgggtttt gtttttttttgcattt gtttttttttgcattt 1920
ttgggttttttgcattt aaagtgttttgcattt gtttttttttgcattt gtttttttttgcattt 1980
aatttttttgcattt tgaatttagat ttttttttttgcattt gtttttttttgcattt gtttttttttgcattt 2040
ttgtgttgcattt ttttttttttgcattt gtttttttttgcattt gtttttttttgcattt 2100
gatgaatttagt ttttttttttgcattt gtttttttttgcattt gtttttttttgcattt 2160
tataaagggtt ttttttttttgcattt gtttttttttgcattt gtttttttttgcattt 2220
```

-continued

tgaaaatttg taaaaagtaa aataatgaga attaggttt aggtgttaggt tgtttagagt	2280
ttaaagtagg tatgttagtag gtattgtgg atatagtgt a tttggagttt gttgaatata	2340
ttaaattttt ttatgtgaat tggatttgaa gaattgtttt tttgaaaga aaaaagatta	2400
ttttgaaat agtaaaaatt atagaaataa aaaagaatag gaattgaatg agaaaaatgt	2460
ttgggggtgg gaggtaaagt ttaatatttt taataattaa aaagttttt tgttgttatt	2520
ttagtatatg aagggttagt ttgtaatattt atagggtgtt atagagtta tttttatgtt	2580
tatggagttt tttttttttt agatattttt ttgtttatattt atttttttt tttttatgtt	2640
tttttttat ttttattttt gtttagttaa ttttaaagtg atgatttaga tattttattt	2700
attgtatttt ttttgttttt ttattttgtt tattttttt atgagaattt gaattttttt	2760
agaatttgtt ttaataattt gtttattaaa gtgaatagaa gatatgaatt tttatggaa	2820
tttataatat taatttgttag gaatagaaaa aaaaaggaga taattttatag tttaaatatta	2880
tttataattt ataattttaa attataattt tagattttaa ggttttatattt aaagaagaaaa	2940
gttattgttt tagtataaga gtgggtttat gtagtggta aaaattttagt attagttttt	3000
t	3001

```

<210> SEQ ID NO 60
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 60
ataaatcatc ccaaaacctc ta

```

22

```

<210> SEQ ID NO 61
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 61
cgcgctactc cgcataca

```

18

```

<210> SEQ ID NO 62
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 62
gaggtaatcg aggccgtcg

```

19

```

<210> SEQ ID NO 63
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 63
cgccaaattca tacgcccac c

```

21

-continued

```
<210> SEQ ID NO 64
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 64
```

```
accgaaaata cgtttcacg 19
```

```
<210> SEQ ID NO 65
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 65
```

```
gcgttatcgt aaagtattgc gc 22
```

```
<210> SEQ ID NO 66
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 66
```

```
cgcgacgaac aaaacgccc 19
```

```
<210> SEQ ID NO 67
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 67
```

```
gcgttttacg tcgtcgcg 18
```

```
<210> SEQ ID NO 68
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 68
```

```
gacgctaaac gccaccgt 18
```

```
<210> SEQ ID NO 69
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 69
```

```
ccgaccatcc gacgccttac tcg 23
```

```
<210> SEQ ID NO 70
<211> LENGTH: 21
```

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 70
```

```
cgttttcgt tttatttcg c 21
```

```
<210> SEQ ID NO 71
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 71
```

```
gacaaaaaac gccacgtc 18
```

```
<210> SEQ ID NO 72
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 72
```

```
ccgacaattc accgaatcac cg 22
```

```
<210> SEQ ID NO 73
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 73
```

```
atctcaccta ccgtcgcg 18
```

```
<210> SEQ ID NO 74
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 74
```

```
taggagtgcg atcggttgc 19
```

```
<210> SEQ ID NO 75
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
<400> SEQUENCE: 75
```

```
acgaacgtta cgaccgatac ccaacta 27
```

1. A method for the detection of cell proliferative disorders, comprising determining the CpG methylation status of two or more of the genes ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences, wherein based on said determined methylation status, the presence or absence a colorectal cell proliferative disorder or metastasis is deduced.

2. The method of claim **1**, wherein said genes are selected from the group consisting of ALX4, TPEF and p16/INK4A.

3. The method of claim **2**, wherein said genes consist ALX4, TPEF and p16/INK4A.

4. A method for the analysis of colorectal cell proliferative disorders, comprising determining the CpG methylation status of the gene ALX4 and/or its regulatory sequences, wherein based on said determined methylation status the presence or absence of a colorectal cell proliferative disorder is deduced.

5. A method for the analysis of colorectal cell proliferative disorders, comprising determining the CpG methylation status of the gene ALX4 and/or its regulatory sequences and one or more of the genes selected from the group consisting TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences, wherein based on said determined methylation status the presence or absence of a colorectal cell proliferative disorder is deduced.

6. A nucleic acid molecule consisting essentially of a sequence at least 18 bases in length according to a sequences selected from the group consisting of SEQ ID NOS: 7,8, 15 and 16.

7. An oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, for the detection of colon cell proliferative disorders, said oligomer consisting essentially of at least one base sequence having a length of at least 10 contiguous nucleotides which hybridises to or is identical to a nucleic acid sequence selected from the group consisting of SEQ ID NOS:7,8, 15 and 16.

8. The method of claim **1**, comprising:

obtaining, from a subject, a biological sample having subject genomic DNA;

contacting the genomic DNA, or a fragment thereof, with one reagent or a plurality of reagents for distinguishing between methylated and non methylated CpG dinucleotide sequences within at least two target sequences of the genomic DNA, or fragment thereof, wherein the target sequence comprises, or hybridizes under stringent conditions to, at least 16 contiguous nucleotides of a sequence taken from the group consisting of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59 said contiguous nucleotides comprising at least one CpG dinucleotide sequence; and

determining, based at least in part on said distinguishing, the methylation state of at least one target CpG dinucleotide sequence, or an average, or a value reflecting an average methylation state of a plurality of target CpG dinucleotide sequences, whereby detecting, or detecting colon cell proliferative disorders is, at least in part, afforded.

9. The method of claim **8**, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises converting unmethylated cytosine bases within the target sequence to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties.

10. The method of claim **8**, wherein the biological sample is selected from the group consisting of cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, stool, blood, and combinations thereof.

11. The method of claim **8**, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises use of at least one nucleic acid molecule or peptide nucleic acid (PNA) molecule comprising, in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59, and complements thereof.

12. The method of claim **1**, comprising:

a) obtaining, from a subject, a biological sample having subject genomic DNA;

b) extracting or otherwise isolating the genomic DNA;

c) treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;

d) contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence of at least 9 nucleotides that is complementary to, or hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59, and complements thereof, wherein the treated genomic DNA or the fragment thereof is either amplified to produce at least one amplicate, or is not amplified; and

e) determining, based on a presence or absence of, or on a property of said amplicate, the methylation state of at least one CpG dinucleotide of a sequence selected from the group consisting SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotides of a sequence selected from the groups consisting of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, whereby at least one of detecting, or detecting and distinguishing between colon cell proliferative disorders is, at least in part, afforded.

13. The method of claim **12**, wherein treating the genomic DNA, or the fragment thereof in c), comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.

14. The method of claim **12**, wherein contacting or amplifying in d) comprises use of at least one method selected from the group consisting of: use of a heat-resistant DNA polymerase as the amplification enzyme; use of a polymerase lacking 5'-3' exonuclease activity; use of a polymerase chain reaction (PCR); generation of a amplicate nucleic acid molecule carrying a detectable labels; and combinations thereof.

15. The method of claim **14**, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case modified at the 5'-end thereof to preclude degradation by an enzyme having 5'-3' exonuclease activity.

16. The method of claim **14**, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case lacking a 3' hydroxyl group.

- 17.** The method of claim **16**, wherein the amplification enzyme is a polymerase lacking 5'-3' exonuclease activity.
- 18.** The method of claim **12**, wherein determining in e) comprises hybridization of at least one nucleic acid molecule or peptide nucleic acid molecule in each case comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59 and complements thereof.
- 19.** The method of claim **18**, further comprising extending at least one such hybridized nucleic acid molecule by at least one nucleotide base.
- 20.** The method of claim **12**, wherein determining in e), comprises sequencing of the amplicate.
- 21.** The method of claim **12**, wherein contacting or amplifying in d), comprises use of methylation-specific primers.
- 22.** A kit comprising a bisulfite reagent as well as at least one oligomer consisting essentially of at least one base sequence having a length of at least 10 contiguous nucleotides which hybridises to or is identical to a nucleic acid sequences selected from the group consisting of SEQ ID NOS:5 to SEQ ID NO:0 and SEQ ID NOS:48 to SEQ ID NO: 59.
- 23.** (canceled)

* * * * *